Discovering history through language:

papers in honour of Malcolm Ross

Pacific Linguistics 605

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.

Discovering history through language:
papers in honour of Malcolm Ross

Edited by Bethwyn Evans

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

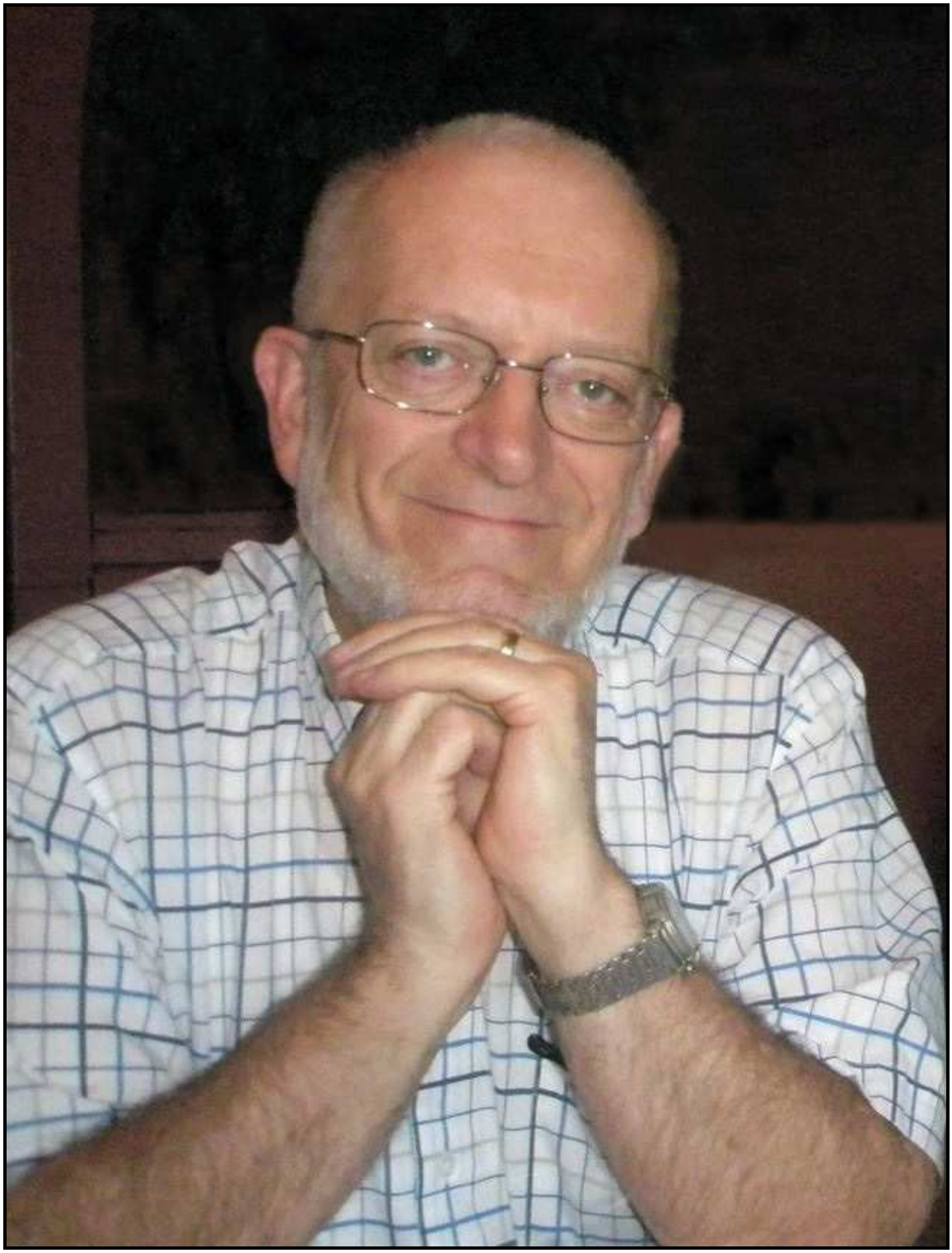
Published by Pacific Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra ACT 0200
Australia

First published 2009

National Library of Australia Cataloguing-in-Publication entry

Title:	Discovering history through language : papers in honour of Malcolm Ross / editor Bethwyn Evans.
ISBN:	9780858836051 (pbk.)
Notes:	Bibliography.
Subjects:	Ross, Malcolm. Historical linguistics. Anthropological linguistics. Linguistic change.
Other Authors/ Contributors:	Evans, Bethwyn.
Dewey Number:	417.7

Copyedited by Bethwyn Evans/Jeanette Coombes
Typeset by Jeanette Coombes
Cover design by Julie Manley
Photo of Lapita pottery fragment courtesy of Christophe Sand, New Caledonia
Printed and bound by Addcolour Digital Pty Ltd, Fyshwick, Canberra



Malcolm Ross

Table of contents

Part II: Historical development of languages across time

Contributors to this volume

I Wayan Arka
Department of Linguistics, RSPAS
The Australian National University
Canberra ACT 0200
Australia
wayan.arka@anu.edu.au

Roger Blench
Kay Williamson Educational Foundation
8 Guest Road
Cambridge CB1 2AL
United Kingdom
rogerblench@yahoo.co.uk

Robert A. Blust
Department of Linguistics
University of Hawai'i at Manoa
1890 East-West Road
Honolulu, HI 96822
USA
blust@hawaii.edu

Mark Donohue
Department of Linguistics, RSPAS
The Australian National University
Canberra ACT 0200
Australia
mark.donohue@anu.edu.au

Robert Early
University of the South Pacific
PMB 9072
Port Vila
Vanuatu
early_r@vanuatu.usp.ac.fj

Bethwyn Evans
Comparative Population Linguistics Group
Max Planck Institute for Evolutionary Linguistics
Deutscher Platz 6
04103 Leipzig
Germany
bethwyn.evans@googlemail.com

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

Alexandre François
LACITO – CNRS
7 rue Guy Môquet
F – 94801 Villejuif
France
Alexandre.Francois@vjf.cnrs.fr

Harold Koch
School of Language Studies
The Australian National University
Canberra ACT 0200
Australia
Harold.Koch@anu.edu.au

Paul Jen-kuei Li
Institute of Linguistics, Academica 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

Anna Margetts
Linguistics Program
Monash University
Victoria 3800
Australia
Anna.Margetts@arts.monash.edu.au

Johanna Nichols
Department of Slavic Languages and
Literatures
#2979 University of California
Berkeley, CA 94720-2979
USA
johanna@berkeley.edu

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

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

Ger Reesink
Centre for Language Studies
Radboud University Nijmegen
Erasmusplein 1 6525 HT Nijmegen
The Netherlands
ger.reesink@hccnet.nl

Paul Sidwell
Department of Linguistics, RSPAS
The Australian National University
Canberra ACT 0200
Australia
paulsidwell@yahoo.com

Jacinta Smallhorn
Department of Linguistics, RSPAS
The Australian National University
Canberra ACT 0200
Australia
jacinta.smallhorn@anu.edu.au

Tonya Stebbins
Linguistics Program
La Trobe University
Victoria 3086
Australia
T.Stebbins@latrobe.edu.au

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

Laurie Reid
5-1-1-1701 Nyoidani
Minoo-shi
Osaka 562-0011
Japan
reid@hawaii.edu

Stacy Teng
Institute of Linguistics, Academica Sinica
128, Sec. 2, Academia Road
11529 Nankang, Taipei
Taiwan
stacyteng@gmail.com

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

Acknowledgements

I write this looking back to COOL7 (7th Conference on Oceanic Linguistics) in Nouméa and forwards to COOL8 in Auckland as they represent the beginning and end of putting together this volume. It was in Nouméa in 2007 that Andrew Pawley first proposed the idea of a festschrift volume for Malcolm, and when it was decided to aim for COOL8 for the volume's launch.

Malcolm's research focus on historical linguistics made it an obvious theme for the volume. It is hoped that he is pleased with this collection of papers, which touch on many aspects of his own research. I would like to thank all the authors for not only making this volume possible, but for contributing such an interesting collection of papers. I am also grateful to the many other scholars who acted as anonymous referees.

Andrew Pawley, Meredith Osmond and Ingrid Ross have shared with me their memories of Malcolm and his research, helping to make the volume one that truly does honour his career as a linguist. I am grateful to Andy for his assistance in planning the volume, editing some of the papers, and for all his useful and timely suggestions along the way. And also to Meredith for her help with editing and formatting many of the papers.

Pacific Linguistics not only agreed to publish this volume, but have made every effort to ensure its timely publication. I am grateful to the managing editors, Wayan Arka, Mark Donohue and John Bowden, for their support, and especially to Jeanette Coombes for copy editing and typesetting. Thanks also to Karina Pelling for drawing the maps and to Ingrid Ross for providing the photos of Malcolm.

Bethwyn Evans
November 2009, Leipzig

1 *Malcolm Ross: an appreciation*

BETHWYN EVANS

The 'loom of language'[1]

Bodmer's (1944) *The Loom of Language* is a volume rarely cited within the historical linguistics literature. Yet, indirectly this book has contributed significantly to the way in which many aspects of historical linguistics are viewed today.

Bodmer sets out for school students some of the foundations of historical linguistics, including the ways in which languages are connected to each other historically, the ways in which their structure changes through time, and the ways in which they spread across time and space. He writes, for example:

> Languages are *related*, if the many features of vocabulary, structure and phonetics which they share are due to the gradual differentiation of what was once a single tongue. Sometimes we have to infer what the common parent was like …
>
> Through such culture-contacts words have wandered from one language to another of a totally different origin. The modern word bicycle pedals over linguistic frontiers as the machine used to pedal over national boundaries before passports were obligatory.
>
> (Bodmer 1944:167, 184)

Perhaps now somewhat dated in expression and lacking in theoretical rigour, *The Loom of Language* remains inspiring. In the 1950s it captured the imagination of a 12-year old boy, planting the seeds for an academic career which centred around historical linguistic research. The research of this boy, now Emeritus Professor Malcolm Ross, has contributed to many domains of historical linguistics, and itself forms the foundation for the research of others.

Young Malcolm

Malcolm Ross often calls himself 'an accidental linguist', not beginning his PhD in linguistics until the age of 40. But Malcolm's career in linguistics was clearly an accident waiting to happen, with his interest and curiosity in language and linguistics apparent from an early age. Born in London in 1942, Malcolm's introduction to languages other than his own was as a schoolboy 'reading' Welsh language newspapers on family holidays and devouring any book relating to language and linguistics he could find in the local library. As a high school student Malcolm had the chance to study other European languages:

[1] I would like to thank Meredith Osmond, Andrew Pawley and Ingrid Ross for commenting on earlier versions of this piece and for sharing with me some of their memories of Malcolm.

> He had French (first) then Latin, then Greek at school. He took Russian in the lunch hour. He then had the option of Spanish as an O-level subject in the 6th form for two years (although he did 3 years in the A-level course. He had done his O-levels in 4 years instead of 5 as he was in the 'Fast-track class'). (Ingrid Ross, pers.comm.)

Student days, Bristol

Without the opportunity to undertake formal study of linguistics, Malcolm went on to study English Literature at Bristol University, writing his BA honours thesis on mediæval political and satirical verse. However, his interest in language remained apparent; he took as many linguistics courses as he could, and:

> he took all the language options like Old English, Middle English and Old Norse. He also took Spanish as a subsidiary subject. (Ingrid Ross, pers.comm.)

And he added a 'smattering' of Portuguese to his repertoire of languages when he went to Portugal – by bicycle.

It was while he was at University that Malcolm met his wife Ingrid, who was also studying at Bristol University as part of an exchange programme with her own university in Germany. Their meeting on a student hike was the beginning of what has been a long and happy partnership. This was also when Malcolm, unsurprisingly, began to learn German, a language in which he has gained near native fluency.

Malcolm the teacher

After University, Malcolm followed Ingrid back to Germany. In 1964–65 he worked at the *Sprachenschule der Stadt Müchen* in Germany, training translators and interpreters in English. This also allowed him to practice his German — 'no-one was allowed to speak to him in English' (Ingrid Ross, pers.comm.). In 1965, after marrying in Koblenz on the Rhine, Malcolm and Ingrid came back to southwest England, and Malcolm became a high-school teacher of English (and some Russian) at Filton High School. He remained at Filton for eight years, ending up the School's Director of Studies (1970–72). While teaching at Filton, he did his Master of Letters in Education and continued learning foreign languages. As Ingrid remembers:

> When he did his M.Litt. in East German Education, he ordered 'Pravda' (probably put him on the CIA black list) and started reading that, but also learnt to read (not speak) the Scandinavian languages and Dutch and the Slavic languages and Hungarian. He used to have a vocab book for each and learn relevant vocab just before he went to sleep and then put the vocab book under his pillow. Seemed to work!

As a student of Malcolm's during his subsequent career in linguistics, I have no doubt that his success and rapid promotion through the teaching ranks was due to his skill and patience in explaining clearly complex ideas and concepts and encouraging students to explore their own ideas. However, like many good teachers, promotion led Malcolm away from teaching and into more and more administrative tasks. In order to 'avoid becoming another school administrator', he decided to change career path, accepting an Education Officer position in Papua New Guinea.

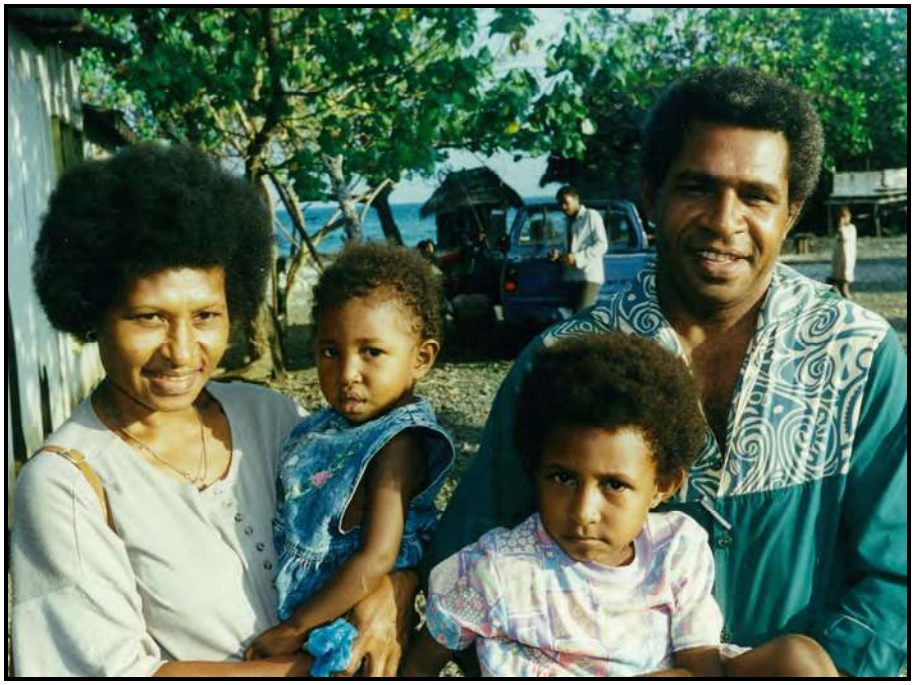
Malcolm in Papua New Guinea

In 1973 Malcolm and Ingrid, along with their two young children, Philippa and James, moved to Papua New Guinea; a rather drastic change of direction career-wise for Malcolm, but Malcolm and Ingrid were looking for a change, and somewhere that would suit young children. An advertisement in the paper for a teaching post near Rabaul on New Britain with a photo of an idyllic Pacific scene — palm trees, beach and swirling seas — looked like just the right combination. And certainly it was an ideal place to be for someone with Malcolm's interest in language.

Summer 1973 at Kerevat

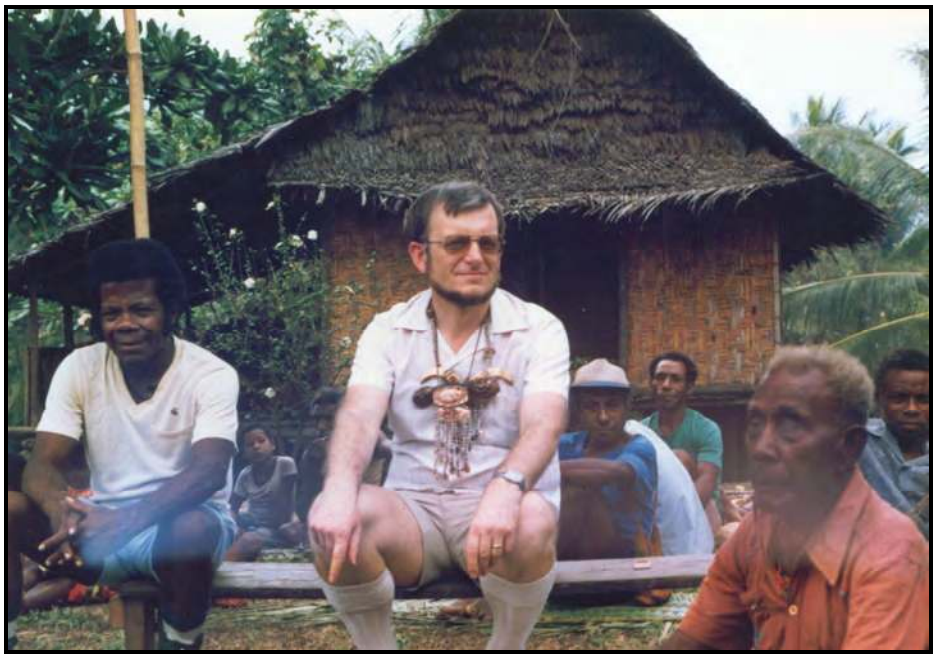
In 1973–74, Malcolm founded the English department at Kerevat National High School, near Rabaul on New Britain. It was on his first day at Kerevat that he began to think about linguistic research. With his characteristic interest in his students, he asked a group of them what they studied in their history classes. Their reply, that they studied Indonesian and Australian history because, as they described it, 'they did not have any history of their own', shocked Malcolm. Certainly there was very little written history of Papua New Guinea, but did that mean that nothing could be known of its past? Surrounded by students speaking many different languages and remembering what he'd learnt from Bodmer on inferring linguistic history, he wondered if language was a way to discover aspects of Papua New Guinea's unwritten past, and so give students some history of their own. And so began Malcolm's interest in Austronesian, and more specifically Oceanic, historical linguistics.

During their stay at Kerevat Malcolm and Ingrid 'adopted' Mait Kilil, a student, who took them to Karkar Island and made them part of his clan, and with whom Malcolm then did extensive collection of linguistic data.



Mait and family, Madang

In 1975 the Ross family moved to Goroka in the Eastern Highlands Province, where Malcolm taught language studies and trained high school teachers at Goroka Teachers' College for several years before becoming Principal in 1980. Trainee teachers came to study in Goroka from all over Papua New Guinea, and once again Malcolm was surrounded by speakers of a plethora of different Austronesian and Papuan languages, many of which had never been recorded. He collected linguistic data, typically basic vocabulary and grammatical structures, from over 150 languages and dialects. His fieldnotes from this time are meticulous — very neat handwritten pages giving consistently-ordered and comparable data for each language, a very accessible resource for those students he lets loose on his filing cabinets.



Malcolm with informants, Karkar Island, 1978

At first Malcolm's research focus was on synchronic analysis, leading to descriptive publications on the Papuan languages Waskia (Ross and Poal 1978) and Vanimo (Ross 1980). However, his interest in Austronesian historical linguistics was growing.

It was during his time at Goroka that Malcolm came to the notice of the academic community. In early 1976 he spent two weeks visiting The Australian National University (ANU) in Canberra in order to learn more about Austronesian linguistics, especially from scholars in the Department of Linguistics in what was then the Research School of Pacific Studies. Whenever Malcolm talks of this first visit to ANU, he comments on discussions with Robert (Bob) Blust, then a Postdoctoral Fellow in the Department, on how much he learnt from Bob at the time, and on how these discussions helped him begin the research that became his PhD thesis. Back in Papua New Guinea, Malcolm not only continued to collect comparative data from Oceanic languages, but he began to research aspects of Oceanic linguistic history. In May 1977, he sent a draft paper on the Oceanic languages of the Vitiaz Straits and the north coast of Papua New Guinea to Andrew Pawley, writing:

> Enclosed is a copy of my first excursion into AN [Austronesian], which was occasioned partly by the fact that I had appropriate informants to hand and partly by reading various pieces of your work, particularly on Eastern Oceanic and Central Papuan, both of which fascinated me. I'm afraid that my efforts probably have rough edges, and any comments would be most welcome.[2]

In response, Andy not only congratulated Malcolm on a formidable paper, but asked 'where did you spring from?' At that time there were only a few scholars active in Oceanic historical linguistics, so one can imagine Andy's surprise at receiving such a paper from someone he had never heard of. Malcolm's reply, telling how he ended up in Papua New Guinea doing research on Oceanic languages, highlights his interest in using linguistic research to discover the past, as well as to assist his students in their role as teachers:

---

[2] Letter from Malcolm Ross to Andrew Pawley, May 1977.

> I've been in PNG since Jan 1973, and became more and more involved in looking at the [languages] here because I thought it important to teach my students (especially those who intend to teach English) something about the sociolinguistic and historical-linguistic situation in PNG. Finally I spent a week at ANU in Jan 1976 trying to fill in gaps and find out, with much help from Bob Blust, what was going on. In Feb I discovered that I had a whole gaggle of AN-speaking students from the Aitape area and decided that the opportunity was too good to miss. I also had an excellent student, John Natu Paol, who had done some research into his own language, Waskia (Karkar and neighbouring mainland at Tokain), so I became involved in that at the same time. Stephen Wurm is going to publish the resulting grammar sketch and vocab. in [Pacific Linguistics].[3]

Graduation ceremony, Goroka, 1981

Malcolm the PhD student

Although Malcolm was carrying out linguistic research, publishing papers, and attending conferences in the late 1970s, he did not begin his PhD until 1982. For this he went back to the ANU and joined Stephen Wurm's Department of Linguistics in the Research School of Pacific Studies, the leading centre for the study of the languages of Papua New Guinea and where he had received a warm welcome in 1976. The primary goal of Malcolm's PhD project, undertaken between 1982 and 1986, was to investigate genetic relationships amongst the more than 250 Austronesian languages of western Melanesia, an area including Papua New Guinea and the northwestern Solomon Islands. All the languages concerned belong to the large Oceanic branch of Austronesian which consists of the languages of eastern Melanesia and Polynesia, and most of the languages of Micronesia.

---

3 Letter from Malcolm Ross to Andrew Pawley, 2nd July 1977.

Malcolm's PhD thesis, published as a monograph (Ross 1988), changed the face of Oceanic linguistics. Described by one examiner, George Grace, as a 'landmark work' it is amongst the most frequently cited works within the field of Oceanic historical linguistics. It was a landmark for a number of different reasons.

First, the thesis was based on the extensive primary data that Malcolm had collected while living in Papua New Guinea. The majority of these data were from languages of western Melanesia which there was little or no previous record. Malcolm's near exhaustive sample brought many languages under the comparative microscope for the first time.

Second, the breadth and depth of Malcolm's analysis was truly epic. He rigorously applied the Comparative Method to more than 200 languages falling into dozens of lower-order subgroups. This systematic and careful comparison, allowed him to determine regular sound correspondences amongst them and to prepare the way for reconstruction of the phonological histories of each subgroup. He did not restrict himself to lexicophonological comparison, but also investigated a range of morphological and syntactic characteristics of the languages.

As the aim of Malcolm's research was to establish subgroupings, he was reliant on determining shared innovations calibrated from the reconstructed Proto Oceanic system. Thus the third contribution of the thesis was to revise and add to the body of work on Proto Oceanic.

Fourth, and very significantly, Malcolm presented a new subgrouping of the Oceanic languages in the western Melanesia region that radically altered the view of the Oceanic family tree. Since Dempwolff (1927, 1937), Oceanic had been widely, though not unanimously, accepted as a major subgroup of the Austronesian language family. However, its internal classification, particularly the languages of western Melanesia, was less certain. When Malcolm began his doctoral studies, only small low-level groupings of languages had been recognised in western Melanesia. For example, Pawley and Green (1985) listed 21 different subgroups of Oceanic languages in Papua New Guinea and the northwestern Solomon Islands. Scholars were also debating the possible historical explanations for the high degree of linguistic diversity found in this region (Pawley 1981; Lynch 1981). On the basis of his detailed comparisons, Malcolm presented evidence to demonstrate that the languages of Papua New Guinea and the northwestern Solomon Islands formed three primary subgroups of Oceanic, namely Western Oceanic, with its three sub-branches, North New Guinea, Meso-Melanesian and Papuan Tip, encompassing most of the region, and the Admiralties and St Matthias subgroups, named after the groups of islands in Papua New Guinea where the languages are spoken. For each of these subgroups, as well as the lower-level groups within them, Malcolm described the phonological, morphological and syntactic innovations that supported them.

Fifth, Malcolm's thesis was also significant in that it redressed a strong and longstanding bias in Oceanic linguistics towards the languages often classified as 'Eastern Oceanic', that is the languages of the southeastern Solomon Islands, Vanuatu, New Caledonia, Micronesia, Fiji and Polynesia. The Oceanic languages which formed part of Dempwolff's evidence for the Austronesian language family and reconstructions of Proto Malayo-Polynesian were all 'Eastern' ones,[4] and until the 1980s the leading scholars in Oceanic historical linguistics, including George Grace, Bruce Biggs, Andrew Pawley, John Lynch, Darrell Tryon and Paul Geraghty, worked predominantly, though not exclusively,

[4] Dempwolff did not include the Formosan languages of Taiwan in the family, and so his reconstructions are now considered to represent the level of Proto Malayo-Polynesian, rather than Proto Austronesian.

on 'Eastern Oceanic' languages. This meant that the understanding of Oceanic languages and their history was based primarily on fewer than half the languages within the subgroup, with comparative research on the languages of western Melanesia lagging far behind that on 'Eastern Oceanic' languages. Happily, this is no longer the case.

Finally, such a meticulous application of the Comparative Method raised a number of theoretical issues regarding the nature of subgrouping. In particular, Malcolm proposed in his thesis a difference between what he later labelled innovation-defined and innovation-linked subgroups. These two types of subgroups represent different patterns of diversification, namely the sharp separation of an original protolanguage into discrete branches versus the gradual differentiation of a network of dialects, and these two types are reflected by different patterns of shared innovations (Ross 1988, 1997).

Malcolm the 'real' linguist

After completing his PhD in 1986, Malcolm stayed on at the ANU. And so in 1987, at the age of 44, he began his career as a professional researcher and teacher in linguistics. Since then he has held a number of different research positions within the Department of Linguistics, Research School of Pacific and Asian Studies (RSPAS), becoming Professor in 2004.

Though he was a late starter, Malcolm has more than made up for it. He has built an outstanding record as both a researcher and a teacher. During his time in RSPAS, he has supervised around 30 PhD students, a task he carries out with a mix of encouragement and constructive criticism, and a genuine interest in his students' research. Malcolm's teaching in linguistics has not been limited to the supervision of graduate students. He has regularly co-taught courses on Austronesian and Papuan languages at the ANU, and has been invited to give similar courses at a number of other institutions, including the J.W. Goethe University, Frankfurt/Main, Christian Albrecht University, Kiel and the University of Stuttgart in Germany, Academica Sinica, the National Taiwan University and the National Tsing Hua University in Taiwan, and at Cornell University and the University of California Berkeley as part of the Linguistic Society of America's Summer Institutes.

Malcolm's research achievements have been both numerous and diverse, but with a focus on discovering history through language, which covers nearly all aspects of historical linguistics. And although his research has centred on Austronesian and Papuan languages, its impact has reached far beyond these spheres.

Reflections of the weave

Working as he does on Austronesian and Papuan languages, which do not have a long written tradition, Malcolm's historical linguistics research relies on comparison of contemporary languages in order to infer the patterns or 'weave' of the linguistic past. As described above, his PhD research was based on his own extensive sample of primary data, and throughout his career he has contributed to the synchronic description of Austronesian and Papuan languages. Prior to his PhD, Malcolm published a number of descriptive studies on languages of the western Melanesia region, including Papuan Waskia (Ross and Paol 1978; Karkar Island, Madang Province) and Vanimo (Ross 1980; West Sepik Province), as well as the Austronesian languages of New Ireland (Ross 1982). Malcolm has continued to publish descriptive research on the Austronesian languages of this region, often based on his original fieldwork from his time in Papua New Guinea; for example, the

14 grammatical sketches authored or co-authored by Malcolm in *The Oceanic Languages* volume (Lynch, Ross and Crowley 2002). While most of his synchronic research is on Oceanic languages, and in particular Takia (Ross 1994b, 2004a), his more recent interest in the Austronesian languages of Taiwan has also resulted in descriptive work on Puyuma, undertaken with his former student Stacy Teng (Ross and Teng 2005; Ross 2008b). Despite Malcolm's clear interest and skill in synchronic description, one always feels that he cannot help but view the contemporary data as the reflection of a linguistic past that is just waiting to be discovered and explained.

Reconstructing the warp

An ever-present theme of Malcolm's research has been what within Bodmer's 'loom of language' metaphor could be described as the warp, that is, the unbroken lengthwise threads which form the basis of the entire weave from its beginning to its end. In this sense, Malcolm has concerned himself with questions about the warp of linguistic history —what are the threads of contemporary languages that allow us to determine their genetic relationships with other languages and to reconstruct aspects of their linguistic history?

Since his 1988 book Malcolm has continued to add to and revise our understanding of the internal relationships of Oceanic languages. For example, in two papers published in 1996 Malcolm demonstrated that the Sarmi and Jayapura languages of West Papua form a single subgroup, which in turn forms part of the Western Oceanic Linkage (Ross 1996c); and that the hard-to-classify Yapese language contains strata from several sources but is fundamentally an Oceanic language, and one that does not form a subgroup with any other Oceanic language (Ross 1996b). Both these papers highlight the typical 'Malcolm approach' — careful analysis of the extant data and testing of the evidence against different possible hypotheses. For the Yapese paper in particular, Malcolm shows how detailed analysis of lexicon and sound correspondences can be used to distinguish the 'warp threads' from other aspects of a language's history. More recently he has shown that languages previously thought by some to be non-Austronesian or Papuan — the Solomon Island languages of Kazukuru in New Georgia (Ross and Dunn 2007), and Äiwoo, and other languages of the Reef and Santa Cruz islands (Ross and Næss 2007) — are in fact Oceanic. Ross and Næss show that that the Reefs-Santa Cruz languages form a subgroup with the languages of Utupua and Vanikoro, and this subgroup, which they call, Temotu, appears to have no close relatives within Oceanic.

Malcolm's research on the genetic relationships of languages is not restricted to the languages of the Oceanic subgroup. He has also worked on non-Oceanic Austronesian languages (e.g. Ross 2004c, 2005b), and has recently set himself the massive task of reviewing in detail the phonological evidence for the much-debated subgrouping of this part of the Austronesian family. So it seems likely that 'Ross' will become equally cited on genetic relations in both the east and west of the Austronesian family.

As part of the *Comparative Papuan Project*, initiated in the mid-1990s by Andrew Pawley, Bill Foley and Malcolm, Malcolm has also done important work on the history of of the 700 or so Papuan languages of Melanesia and eastern Indonesia. In this context the initial focus of his research was on establishing genetic relatedness amongst languages, rather than on subgrouping of languages known to be related. The question of what are reliable diagnostics of relatedness amongst languages is still debated in the literature. For Papuan languages, Malcolm has taken a practical approach that is theoretically grounded. In a number of papers (Ross 1995b, 2001b, 2005a), he argues that comparison of pronoun

paradigms across Papuan languages provides a powerful tool for detecting genetic groupings, and that these comparisons are a fairly reliably indication of the extent of the large Trans New Guinea family and the existence of 22 other Papuan language families. He shows that the pronominal forms supporting the Trans New Guinea family are statistically unlikely to result from chance, and are also unlikely to be the result of borrowing when compared with the sociolinguistic contexts in which pronouns have been borrowed elsewhere (Ross 2005a). Malcolm does not dismiss the traditional and well-accepted methods of historical linguistics here; rather he uses pronominal paradigms as initial evidence of genetic relatedness, which can then be tested by applying the Comparative Method. And indeed Andrew Pawley has shown that the Trans New Guinea family is supported by a fairly extensive set of lexical cognates, a reconstructable protophonology and reconstructable aspects of verb morphology (Pawley 2001, 2005).

Malcolm's interest in the 'warp' of linguistic history is not limited to genetic relationships. He has also published extensively, over 40 articles, on linguistic reconstruction. His research on the nature of Proto Oceanic encompasses all aspects of linguistic structure, including: (a) details of its phonological system (e.g. Ross 1989); (b) aspects of morphology and grammar from possession (Ross 2001c) to adjectival categories (Ross 1998a) and spatial deixis (Ross 2003c, 2004b); and (c) lexicon and semantic domains as diverse as meteorological phenomena (Ross 2003b), household artefacts (Osmond and Ross 1998) and staple food crops (Ross 2008c). Again, he has not restricted his research to Oceanic, but has published reconstructions of the Proto Austronesian systems of phonology (Ross 1992), verbal morphology (Ross 1995a) and case marking and personal pronouns (Ross 2006).

While most, if not all, of Malcolm's papers on reconstruction build on Austronesian examples, his concern with general methods of reconstruction and underlying processes of language change is apparent and often made explicit. The edited volume by Mark Durie and Malcolm (1996), which considers the role of the Comparative Method in contexts of regular and irregular language change highlights this concern with methodology and theory.

Reconstructing the weft

Malcolm's 1988 book greatly enhanced our understanding of the genetic relationships amongst the Oceanic languages in western Melanesia. However, it also drew his attention to the considerable amount of data from these languages which could not be accounted for in terms of 'internal' change from features inherited from Proto Oceanic. He concluded that language contact had played an important role in the history of many of these languages. The question of how to explain in more detail the apparently atypical features of these Oceanic languages was the starting point for Malcolm's research on contact-induced language change, and thus the 'weft' of linguistic history — the crosswise, and possibly varying threads, that are interwoven with the warp threads .

Malcolm's initial work on language contact was empirical; investigating two very different instances of contact-induced change in Oceanic languages of western Melanesia. Madak, of New Ireland, Papua New Guinea, has phonological features atypical of an Oceanic language, but which, as Malcolm shows, bear a striking resemblance to the neighbouring Papuan language Kuot. Malcolm argues that such a situation is best explained historically in terms of language shift — speakers of Kuot, or another, now extinct, Papuan language have shifted to an Oceanic language, but speak it with a 'Papuan'

accent (Ross 1994a). Malcolm's other case study of contact-induced change, and the one for which he is most well-known, is that of Oceanic Takia and Papuan Waskia, both spoken on Karkar Island, Madang Province, Papua New Guinea (e.g. Ross 1987, 1996a, 2008a). Takia has many linguistic features that are unusual for an Oceanic language and are different from other Oceanic languages closely related to it. However, these features closely resemble structures found in Waskia. In Takia there has been minimal borrowing of Waskia lexical or grammatical forms, but considerable change towards semantic, lexical and grammatical structures that are the same or similar to ones found in Waskia.

This apparent history of change in Takia challenged models of contact-induced change, especially some of the hypotheses regarding correlations between degrees of social contact and kinds of change (eg. Thomason and Kaufman 1988). Within the field of contact linguistics, Malcolm's name is most frequently associated with the term *metatypy*, which Malcolm coined to denote the kind of structural contact-induced change found in Takia, namely the 'diachronic process' such that

> the morphosyntactic constructions of one of the languages of a bilingual speech community are restructured on the model of the constructions of the speakers' other language (Ross 2007:116).

In formulating his concept of metatypy, Malcolm compared the kinds of contact-induced change that had occurred in the history of Takia with similar changes in other languages (e.g. Kupwar Urdu; Gumperz and Wilson 1971) and drew on the work of George Grace (1981) relating to the calquing of constructions.

The impact of Malcolm's work on metatypic change, is much more significant than this precise characterisation of one kind of contact-induced change implies. In a series of articles (Ross 1987, 1996a, 2001a, 2007, 2008a), Malcolm uses his empirical study of the Takia-Waskia case not only to develop and refine the concept of metatypy, but to discuss its place within a more general model of contact-induced change which aims to better understand the mechanisms of change, as well as its possible psycholinguistic motivations and sociolinguistic contexts. For example, in his 2007 article, he explores in detail the different kinds of contact-induced structural change, including lexical calquing, grammatical calquing and metatypy. Such fine-grained differentiation of kinds of change allows him to investigate the interdependence of different processes of change; concluding that 'metatypy never occurs without calquing' (Ross 2007:139). And from this research Malcolm proposes a strong hypothesis — that 'contact-induced morphosyntactic change in one of a bilingual community's languages entails the restructuring of the whole grammar on the model of the community's other language' (Ross 2007:135) — to be tested with further empirical and theoretical studies as a way towards a greater understanding of the role that contact-induced change plays in language history.

Discovering the pattern of language history

Malcolm has never really separated the 'warp' and the 'weft' of language history, but rather takes all evidence from the contemporary languages and all processes of language change as relevant for reconstructing as accurate a picture as possible of the linguistic past.

Malcolm's research strongly reflects his aim of using language as a window into the past. His models of language differentiation and reintegration, and of contact-induced change (e.g. Ross 1997, 2003a) — based on the careful analysis of linguistic data — he interprets as representing actual events in the past. For example, in his 1997 article, 'Social

networks and kinds of speech communities', Malcolm interprets different patterns of shared innovations as evidence of an 'event' within some original speech community; the gradual weakening of social ties across a speech community (*linkage breaking*) versus a sharp reduction of social interaction between parts of a speech community (*language fissure*) or the gradual versus abrupt (re-)establishing of social networks amongst speakers of two originally distinct speech communities (*linkage rejoining* versus *language fusion*). Malcolm uses the same 'social network' model of speech communities (cf. Milroy and Milroy 1985) to investigate the past social contexts which certain types of contact-induced change are likely to reflect; lexical calquing and metatypic change being likely to reflect an original speech community that had both external social ties and tightknit internal social networks (Ross 2003a, cf. Andersen 1988).

Such a view also paves the way for comparison of the results of historical linguistics with that of other historical disciplines. This is something Malcolm has demonstrated: (a) through the correlation of fine-grained linguistic analysis of shared innovations and their relative sequencing in a group of closely related languages with the absolute dates of events evident in the archaeological record (Ross 1998b); and (b) on a broader scale of Austronesian culture history more generally through the comparison of the results of linguistics, archaeological and population genetic studies (e.g. Pawley and Ross 1993; Sanchez-Mazas et al. 2008; Bellwood et al. forthcoming)

The most monumental product of Malcolm's efforts to discover history through language will undoubtedly be the *Oceanic Lexicon Project*, directed by Malcolm and Andrew Pawley, with Meredith Osmond as research assistant. This project was begun in the early 1990s with the aim of making reconstructions of Proto Oceanic lexicon for a wide range of semantic domains as a way of gaining an understanding of Proto Oceanic speakers' material culture, social organisation, cosmology and categorisation of the physical environment. As part of the first three volumes (seven are planned, in all), Malcolm has authored and co-authored 17 chapters dealing with different semantic domains, including household artefacts, acts of impact, force and change of state, metereological phenomena, properties of inanimate objects, time, and wild and cultivated plants (see Ross, Pawley and Osmond 1998, 2003, 2008). The *Oceanic Lexicon Project* is first and foremost a linguistic project based on careful reconstruction of lexical forms and their meanings using data from as many contemporary languages as possible to put together cognate sets which adhere to established sound correspondences — often made all the more rigorous by Malcolm's 'eagle eye' at spotting forms which do not quite fit the sound correspondences and need further explanation. The project's results clearly contribute to an interdisciplinary approach to the history of Oceania, sometimes supporting and sometimes challenging the results of other disciplines, but always adding to our understanding of the history of the region.

Malcolm in retirement

In 2007, Malcolm retired from his position in RSPAS, but certainly not from linguistics. He remains in RSPAS as Emeritus Professor, continuing with his research and supervision of PhD students — the main difference appearing to be a change in office. A glance at the number of recent and forthcoming articles amongst Malcolm's publications illustrates his continued research productivity. And he is as much in demand as ever to teach and deliver papers abroad — spending five months as Visiting Professor at Academica Sinica, National Tsing Hua University and National Taiwan University in Taipei in 2008–09, and

teaching a six-week course on *Austronesian and Papuan historical linguistics* at the Linguistic Society of America's Summer Institute for 2009 at the University of California at Berkeley. At Berkeley Malcolm also gave the prestigious Hermann and Klara H. Collitz Institute lecture, where, with his typical style of using carefully analysed empirical data to demonstrate a theoretical point, he discussed Oceanic possessive constructions and their implications for mechanisms of language change.

It seems unlikely that Malcolm's research interests and output are going to diminish anytime soon. Not only does he seem to be taking on new projects, he continues on-going projects, like the *Oceanic Lexicon Project*, which with four volumes yet to be published is likely to keep Malcolm, as well as Andy and Meredith, busy for some time to come.

In appreciation of Malcolm

This volume is offered to Malcolm as a token of appreciation from myself, the volume's contributors, and many other colleagues and friends of Malcolm's. Malcolm is a supportive and encouraging mentor and a wise and generous colleague and friend, whose research is both an inspiration and a challenge for the research endeavours of others. And while wishing Malcolm, and Ingrid, all the best for their 'retirement years', we also look forward to Malcolm's continued discoveries on the history of Austronesian and Papuan languages.

References

Andersen, Henning. 1988. Centre and periphery: adoption, diffusion and spread. In Jacek Fisiak, ed. *Historical dialectology*, 39–85. Berlin: Mouton de Gruyter.

Peter Bellwood, Malcolm Ross, Geoff Chambers and Hung Hsiao-chun. forthcoming. Are 'cultures' inherited? Multidisciplinary perspectives on the origins and migrations of Austronesian-speaking peoples prior to 1000 BC. In Ben Roberts and Marc Vander Linden, eds *Investigating archaeological cultures: material culture, variability and transmission*. Dordrecht: Springer.

Bodmer, Frederick. 1944. *The loom of language. A guide to foreign languages for the home student*. London: George Allen and Unwin Ltd.

Dempwolff, Otto. 1927. Das austronesische Sprachgut in den melanesischen Sprachen. *Folia Ethnoglossia* 3.32–43.

—— *Vergleichende Lautlehre des austronesischen Wortschatzes. Band 2: Deduktiv Anwendung des Urindonesischen auf austronesischen Einzelsprachen*. Beihefte zur ZES 17. Berlin: Dietrich Reimer.

Dunn, Michael and Malcolm Ross. 2007. Is Kazukuru really non-Austronesian? *Oceanic Linguistics* 46.210–231.

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

Geraghty, Paul A. 1983. *The history of the Fijian languages*. Honolulu: University of Hawaii Press.

Grace, George. 1981. *An essay on language.* Columbia, South Carolina: Hornbeam Press.

Lynch, John. 1981. Melanesian diversity and Polynesian homogeneity: the other side of the coin. *Oceanic Linguistics* 20.95-129.

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

Milroy, James and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21. 339–384.

Osmond, Meredith and Malcolm Ross. 1998. Household artefacts. In Ross, Pawley and Osmond, eds, 1998, 67–114.

Pawley, Andrew. 1981. Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 269–309. Auckland: Linguistic Society of New Zealand.

—— 2001. The Proto Trans New Guinea obstruents. In Pawley, Ross and Tryon, eds, 2001, 261–300.

—— 2005. The chequered career of the Trans New Guinea hypothesis: recent research and its implications. In Pawley, et al., eds, 2005, 67–107.

Pawley, Andrew, Robert Attenborough, Jack Golson, Robin Hide, eds. 2005. *Papuan pasts: cultural, linguistic and biological histories of the Papuan-speaking peoples*. Canberra: Pacific Linguistics.

Pawley, Andrew and Roger Green. 1985.The Proto Oceanic language community. In Robert Kirk and Emöke Szathmary, eds *Out of Asia: peopling the Americas and the Pacific*, 161–184. Canberra: The Journal of Pacific History.

Pawley, Andrew and Malcolm Ross. 1993. Austronesian historical linguistics and culture history. *Annual review of anthropology* 22.425–459.

Pawley, Andrew, Malcolm Ross and Darrell Tryon, eds. 2001. *The boy from Bundaberg: essays in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics.

Ross, Malcolm. 1980. Some elements of Vanimo, a New Guinea tone language. In *Papers in New Guinea linguistics 20*, 77–109. Canberra: Pacific Linguistics.

—— 1982. Aspect-marking in New Ireland: towards a historical reconstruction. In Rainer Carle, Martine Heinschke, Peter W. Pink, Christel Rost and Karen Stadtlander, eds *Gava': studies in Austronesian languages and cultures dedicated to Hans Kähler*, 173–196. Berlin: Dietrich Reimer.

—— 1987. A contact-induced morphosyntactic change in the Bel languages of Papua New Guinea. In Donald C. Laycock and Werner Winter, eds *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday*, 583–601. Canberra: Pacific Linguistics.

—— 1988. *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.

—— 1989. Proto Oceanic consonant grade and Milke's **nj*. In Ray Harlow and Robin Hooper, eds *VICAL 1: Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics,* 433–495. Auckland: New Zealand Linguistic Society.

—— 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31.23–64.

—— 1994a. Areal phonological features in north central New Ireland. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world*, 551–572. Berlin: Mouton de Gruyter.

—— 1994b. Describing inter-clausal relations in Takia. In Ger P. Reesink ed. *Topics in descriptive Austronesian linguistics*. *Semaian* 11, 40–85. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

—— 1995a. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang and Cheng-hwa Tsang, eds *Austronesian studies relating to Taiwan*, 727–791. Taipei: Institute of History and Philology, Academia Sinica.

—— 1995b. The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world: liber amicorum Bert Voorhoeve,* 139–168. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

—— 1996a. Contact-induced change and the comparative method: cases from Papua New Guinea. In Durie and Ross, eds, 1996, 180–217.

—— 1996b. Is Yapese Oceanic? In Bernd Nothofer, ed. *Reconstruction, classification, description. Festschrift in honor of Isidore Dyen*, 121–166. Hamburg: Verlag Meyer and Co.

—— 1996c. On the genetic affiliations of the Oceanic languages of Irian Jaya. *Oceanic Linguistics* 35.258–271.

—— 1997 Social networks and kinds of speech community event. In Roger M. Blench and Matthew Spriggs, eds *Archaeology and language 1*: *Theoretical and methodological orientations*, 209–261. London: Routledge.

—— 1998a. Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37.85–119.

—— 1998b. Sequencing and dating linguistic events in Oceania: the linguistics/ archaeology interface. In Roger M. Blench and Matthew Spriggs, eds *Archaeology and language. 2: Correlating archaeological and linguistic hypotheses,* 141–173. London: Routledge.

—— 2001a. Contact-induced change in Oceanic languages in north-west Melanesia. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 134–166. Oxford: Oxford University Press.

—— 2001b. Is there an East Papuan phylum? Evidence from pronouns. In Pawley, Ross and Tryon, eds, 2001, 301–321.

—— 2001c. Proto Oceanic **i, *qi* and **-ki.* In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender,* 259–278. Canberra: Pacific Linguistics.

—— 2003a. Diagnosing contact-induced change. In Raymond Hickey, ed. *Motives for language change,* 174–198. Cambridge: Cambridge University Press.

—— 2003b. Meteorological phenomena. In Ross, Pawley and Osmond, eds, 2003*,* 115–148.

—— 2003c. Talking about space: terms of location and direction. In Ross, Pawley and Osmond, eds, 2003, 221–284.

—— 2004a. Aspects of deixis in Takia. In Senft, ed., 2004, 15–36.

—— 2004b. Demonstratives, local nouns and directionals in Oceanic languages: a diachronic perspective. In Senft, ed., 2004, 175–204.

—— 2004c. Notes on the prehistory and internal subgrouping of Malayic. In John Bowden and Nikolaus P. Himmelmann, eds *Papers in Austronesian subgrouping and dialectology*, 97–109. Canberra: Pacific Linguistics.

—— 2005a. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley, et al., eds, 2005, 15–65.

—— 2005b. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2).1–24.

—— 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Henry Y. Chang and 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*, 521–564. Taipei: Institute of Linguistics, Academia Sinica.

—— 2007. Calquing and metatypy. *Journal of Language Contact,* Thema 1.116–143. [Available online http://www.jlc-journal.org/]

—— 2008a. A history of metatypy in the Bel languages. *Journal of Language Contact:* Thema 2.149–164. [Available online http://www.jlc-journal.org/]

—— 2008b. Negative verbal clause constructions in Puyuma: exploring constructional disharmony. In Alexander Bergs and Gabriele Diewald, eds *Constructions and language change,* 171–193. Berlin: Mouton de Gruyter.

—— 2008c. Staple foods: root crops, bananas, breadfruit and sago. In Ross, Pawley and Osmond, eds, 2008*,* 255–292.

Ross, Malcolm and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46(2).456-498.

Malcolm Ross with John Natu Paol. 1978. *A Waskia grammar sketch and vocabulary*. Canberra: Pacific Linguistics.

Malcolm Ross, Andrew Pawley and Meredith Osmond, eds. 1998 *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 1: Material culture*. Canberra: Pacific Linguistics.

—— 2003. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world.* Canberra: Pacific Linguistics.

—— 2008. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants.* Canberra: Pacific Linguistics.

Ross, Malcolm and Stacy Fang-ching Teng. 2005. Clause constructions in Nanwang Puyuma. *Concentric: studies in linguistics* 31.119–158.

Sanchez-Mazas, Alicia, Roger Blench, Malcolm D. Ross, Ilia Peiros and Marie Lin, eds. 2008. *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*. London: Routledge.

Senft, Gunter, ed. 2004. *Deixis and demonstratives in Oceanic languages*. Canberra: Pacific Linguistics.

Thomason, Sarah G. and Terence S. Kaufmann. 1988. *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.

2 *Malcolm Ross – a bibliography*

1 Language in the past[1]

1.1 Genetic relationships amongst languages

1.1.1 Austronesian

1977 *Relationships of the Austronesian languages of the Sepik and western Madang coast of New Guinea. Unpublished manuscript. University of Papua New Guinea.

1979 *The Austronesian languages of Papua: towards a family tree. Unpublished manuscript. University of Papua New Guinea.

1983 *The genetic relationships of the Austronesian languages of Papua. Paper presented to the Austronesian Symposium of the Fifteenth Pacific Science Congress. Dunedin, New Zealand.

1986 A genetic classification of Oceanic languages in Bougainville and the western Solomons. In Paul Geraghty, Lois 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 Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.

1992 The position of Gumawana among the languages of the Papuan Tip cluster. *Language and Linguistics in Melanesia* 23.139–165.

1995 Some current issues in Austronesian linguistics. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary,* Part 1, 45–120. Berlin: Mouton de Gruyter.

1996 On the genetic affiliations of the Oceanic languages of Irian Jaya. *Oceanic Linguistics* 35.258–271.

1996 Is Yapese Oceanic? In Bernd Nothofer, ed. *Reconstruction, classification, description. Festschrift in honor of Isidore Dyen*, 121–166. Hamburg: Verlag Meyer and Co.

2004 Notes on the prehistory and internal subgrouping of Malayic. In John Bowden and Nikolaus P. Himmelmann, eds *Papers in Austronesian subgrouping and dialectology*, 97–109. Canberra: Pacific Linguistics.

[1] * marks unpublished papers.

2005 The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2).1–24.

2006 Language families and linguistic diversity. In Keith Brown, ed. *Encyclopedia of language and linguistics.* 2nd edition, vol. 6, 499–508. Oxford: Elsevier.

2007 [Michael Dunn and Malcolm Ross] Is Kazukuru really non-Austronesian? *Oceanic Linguistics* 46.210–231.

2007 [Malcolm Ross and Åshild Næss] An Oceanic Origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.456–498.

2008 The integrity of the Austronesian language family: from Taiwan to Oceania. In Alicia Sanchez-Mazas et al., eds *Past human migrations in East Asia: matching archaeology, linguistics and genetics,* 161–181. London: Routledge.

In press Lexical history in the Northwest Solomonic languages. In John Bowden and Nikolaus Himmelmann, eds *A journey through Austronesian and Papuan linguistic and cultural space: papers in honour of Andrew K. Pawley.* Canberra: Pacific Linguistics.

1.1.2 Papuan

1995 The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world: liber amicorum Bert Voorhoeve,* 139–168. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

2000 *[Meredith Osmond, Andrew Pawley and Malcolm Ross] Membership and subgrouping of the Trans New Guinea phylum. In Andrew Pawley, Malcolm Ross and Meredith Osmond, eds *Papuan languages and the Trans New Guinea phylum*. Unpublished manuscript. Canberra: RSPAS, The Australian National University.

2000 *Defining the Trans New Guinea family: preliminary evidence from pronouns. In Andrew Pawley, Malcolm Ross and Meredith Osmond, eds *Papuan languages and the Trans New Guinea phylum*. Unpublished manuscript. Canberra: RSPAS, The Australian National University.

2000 *Pronouns as preliminary evidence for grouping Papuan languages. In Andrew Pawley, Malcolm Ross and Meredith Osmond, eds *Papuan languages and the Trans New Guinea phylum*. Unpublished manuscript. Canberra: RSPAS, The Australian National University.

2001 Is there an East Papuan phylum? Evidence from pronouns. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: studies in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

2005 Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew Pawley, Robert Attenborough, Robin Hide and Jack Golson, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Pacific Linguistics.

1.2 Contact relationships amongst languages

1987 A contact-induced morphosyntactic change in the Bel languages of Papua New Guinea. In Donald C. Laycock and Werner Winter, eds *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday*, 583–601. Canberra: Pacific Linguistics.

1991 Refining Guy's sociolinguistic types of language change. *Diachronica* 8.119–129.

1993 The sources of Austronesian lexical items in Tok Pisin. In Tom Dutton, Malcolm Ross and Darrell Tryon, eds *The language game: Papers in memory of Donald C. Laycock*, 361–384. Canberra: Pacific Linguistics.

1994 Areal phonological features in north central New Ireland. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world*, 551–572. Berlin: Mouton de Gruyter.

1996 Contact-induced change and the Comparative Method: cases from Papua New Guinea. In Mark Durie and Malcolm 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 event. In Roger M. Blench and Matthew Spriggs, eds *Archaeology and language I*: *theoretical and methodological orientations*, 209–261. London: Routledge.

1999 *Exploring metatypy: How does contact-induced typological change come about? Keynote address to the Australian Linguistic Society Conference, Perth. [Available online: http://rspas.anu.edu.au/linguistics/projects/mdr/Metatypy.pdf]

2001 Contact-induced change in Oceanic languages in north-west Melanesia. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 134–166. Oxford: Oxford University Press.

2003 Diagnosing contact-induced change. In Raymond Hickey, ed. *Motives for language change,* 174–198. Cambridge: Cambridge University Press.

2006 Metatypy. In Keith Brown, ed. *Encyclopedia of language and linguistics.* 2nd edition, vol. 8, 95–99. Oxford: Elsevier.

2006 *Language contact and language decay: reflections from Papua New Guinea. Paper presented at Third Oxford-Kobe Linguistics Seminar: 'The Linguistics of Endangered Languages'. Kobe, Japan. [Available online http://rspas.anu.edu.au/linguistics/projects/mdr/Kobe_endangerment.pdf]

2007 Calquing and metatypy. *Journal of Language Contact, Thema* 1.116–143. [Available online: http://www.jlc-journal.org/]

2008 A history of metatypy in the Bel languages. *Journal of Language Contact: Thema* 2.149–164. [Available online: http://www.jlc-journal.org/]

1.3 Reconstruction[2]

1.3.1 Austronesian

1979 *Reconstructing Proto-Central Papuan. Unpublished manuscript. University of Papua New Guinea.

1981 *Proto Central Papuan morphosyntax. Unpublished manuscript. Canberra: The Australian National University.

1981 *Proto-Oceanic from the top down. Unpublished manuscript. University of Papua New Guinea.

1982 The development of the verb phrase in the Oceanic languages of the Bougainville region. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.1: *Currents in Oceanic*, 1–57. Canberra: Pacific Linguistics.

1989 Proto Oceanic consonant grade and Milke's **nj*. In Ray Harlow and Robin Hooper, eds *VICAL 1: Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics,* 433–495. Auckland: New Zealand Linguistic Society.

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

1993 Tonogenesis in the North Huon Gulf chain. In Jerold A. Edmondson and Kenneth J. Gregerson, eds *Tonality in Austronesian languages*. Oceanic Linguistics Special Publication. No. 24, 133–153. Honolulu: University of Hawai'i Press.

1995 [R. David Zorc and Malcolm Ross] A glossary of Austronesian reconstructions. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary,* Part 1, Fascicle 2, 1106–1197. Berlin: Mouton de Gruyter.

1995 Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang and Cheng-hwa Tsang, eds *Austronesian studies relating to Taiwan*, 727–791. Taipei: Institute of History and Philology, Academia Sinica.

1995 Proto Oceanic terms for meteorological phenomena. *Oceanic Linguistics* 34.261–304.

1996 [Compiled by Malcolm Ross] On the origin of the term 'Malayo-Polynesian'. *Oceanic Linguistics* 35.143–145.

1996 Pottery terms in Proto Oceanic. In Janet Davidson, Geoffrey Irwin, Foss Leach, Andrew Pawley, and Dorothy Brown, eds *Oceanic culture history: essays in honour of Roger Green,* 67–82. Wellington: New Zealand Journal of Archaeology.

1996 Reconstructing food plant terms and associated terminologies in Proto Oceanic. In John Lynch and Fa'afo Pat, eds *Oceanic studies: proceedings of the First International Conference on Oceanic Linguistics*, 163–221. Canberra: Pacific Linguistics.

---

[2] Many listed in §1.1 may be included here.

1996 [Malcolm Ross and Mark Durie] Introduction. In Mark Durie and Malcolm Ross, eds *The Comparative Method reviewed: regularity and irregularity in language change,* 3–38. New York: Oxford University Press.

1998 Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37.85–119.

1998 Possessive-like attribute constructions in the Oceanic languages of northwest Melanesia. *Oceanic Linguistics* 37.234–276.

1998 [Malcolm Ross, Andrew Pawley, and Meredith Osmond] Introduction. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society*. *1: Material culture*, 1–14. Canberra: Pacific Linguistics.

1998 Proto Oceanic phonology and morphology. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society*. *1: Material culture*, 15–35. Canberra: Pacific Linguistics.

1998 [Meredith Osmond and Malcolm Ross] Household artefacts. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 1*: *Material culture*, 67–114. Canberra: Pacific Linguistics.

1998 Acts of impact, force and change of state. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society*. *1: Material culture*, 233–295. Canberra: Pacific Linguistics.

2000 Proto Oceanic adjectival morphology: the suffix *-[k]a.* In Steven Roger Fischer and Wolfgang B. Sperlich, eds *Leo Pasifika: Proceedings of the Fourth International Conference on Oceanic Linguistics*, 326–342. Auckland: The Institute of Polynesian Languages and Literatures.

2001 Proto Oceanic **i, *qi* and *-ki.* In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology: a focusschrift for Byron W. Bender,* 259–278. Canberra: Pacific Linguistics.

2001 [Bethwyn Evans and Malcolm Ross] The history of Proto-Oceanic *ma-. *Oceanic Linguistics* 40.269–290.

2002 The history and transitivity of western Austronesian voice and voice-marking. In Fay Wouk and Malcolm Ross, eds *The historical and typological development of western Austronesian voice systems,* 17–62 . Canberra: Pacific Linguistics.

2002 Final words: research themes in the history and typology of western Austronesian languages. In Fay Wouk and Malcolm Ross, eds *The historical and typological development of western Austronesian voice systems,* 451–474. Canberra: Pacific Linguistics.

2003 [Meredith Osmond, Andrew Pawley and Malcolm Ross] Landscape. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world.* 35–86. Canberra: Pacific Linguistics.

2003 [Meredith Osmond, Andrew Pawley and Malcolm Ross] Seascape. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world.* 87–113. Canberra: Pacific Linguistics.

2003 Meteorological phenomena. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world,* 115–148. Canberra: Pacific Linguistics.

2003 Properties of inanimate objects. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world*, 187–220. Canberra: Pacific Linguistics.

2003 Talking about space: terms of location and direction. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world,* 221–284. Canberra: Pacific Linguistics.

2003 Time. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world*, 285–325. Canberra: Pacific Linguistics.

2004 Demonstratives, local nouns and directionals in Oceanic languages: a diachronic perspective. In: Gunter Senft, ed. *Deixis and demonstratives in Oceanic languages*, 175–204. Canberra: Pacific Linguistics.

2005 The morphology of some Oceanic plant names. In Claudia Gross, Harriet D. Lyons and Dorothy A. Counts, eds *A polymath anthropologist: essays in honour of Ann Chowning,* 197–204. Auckland: Department of Anthropology, University of Auckland.

2006 Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Henry Y. Chang and 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,* 521–564. Taipei: Institute of Linguistics, Academia Sinica.

2008 Wild plants of the coastal strand. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 129–172. Canberra: Pacific Linguistics.

2008 Wild plants of the mangrove swamp. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 173–183. Canberra: Pacific Linguistics.

2008 Wild plants of primary lowland tropical rain forest. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 185–236. Canberra: Pacific Linguistics.

2008 Wild plants of secondary lowland rain forest and grassland. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 237–254. Canberra: Pacific Linguistics.

2008 Staple foods: root crops, bananas, breadfruit and sago. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 255–292. Canberra: Pacific Linguistics.

2008 Green vegetables and figs. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 293–310. Canberra: Pacific Linguistics.

2008 Nut and fruit trees. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 311–353. Canberra: Pacific Linguistics.

2008 [Malcolm Ross and Bethwyn Evans] The coconut palm. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 355–387. Canberra: Pacific Linguistics.

2008 Other cultivated plants. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 389–426. Canberra: Pacific Linguistics.

2008 Concluding notes. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants,* 427–436. Canberra: Pacific Linguistics.

2009 Proto Austronesian verbal morphology: a reappraisal. In Alexander Adelaar and Andrew Pawley, eds *Austronesian historical linguistics and culture history: a festschrift for Robert Blust*, 295–326. Canberra: Pacific Linguistics.

1.3.2 Papuan

2005 Towards a reconstruction of the history of tone in the Trans New Guinea family. In Shigeki Kaji, ed. *Proceedings of the Symposium Crosslinguistic Studies of Tonal Phenomena: historical development, tone-syntax interface, and descriptive studies,* 3–31. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

1.4 Linguistics in culture history

1989 Early Oceanic linguistic prehistory: a reassessment. *Journal of Pacific History* 24. 135–149.

1991 How conservative are sedentary languages? Evidence from western Melanesia. In Robert Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace,* 443–451. Canberra: Pacific Linguistics.

1993 [Andrew Pawley and Malcolm Ross] Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22.425–459.

1994 Central Papuan culture history: some lexical evidence. In Andrew Pawley and Malcolm Ross, eds *Austronesian terminologies: continuity and change*, 389–479. Canberra: Pacific Linguistics.

1995 [Andrew Pawley and Malcolm Ross] The prehistory of Oceanic languages: a current view. In Peter Bellwood, James Fox and Darrell Tryon, eds *The Austronesians: historical and comparative perspectives*, 39–74. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, The Australian National University.

1997 Comment on 'The dimensions of social life in the Pacific' by J.E. Terrell, T.L. Hunt and C. Gosden. *Current Anthropology* 38.182–184.

1998 Sequencing and dating linguistic events in Oceania: the linguistics/archaeology interface. In Roger M. Blench and Matthew Spriggs, eds, *Archaeology and language. II: Correlating archaeological and linguistic hypotheses,* 141–173. London: Routledge.

2008 [Roger Blench, Malcolm Ross and Alicia Sanchez-Mazas] Methodological issues: linking genetic, linguistic and archaeological evidence. In Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross, Ilia Peiros and Marie Lin, eds *Past human migrations in East Asia: matching archaeology, linguistics and genetics,* 3–19. London: Routledge.

Forthcoming. [Peter Bellwood, Malcolm Ross, Geoff Chambers and Hung Hsiao-chun]. Are 'Cultures' inherited? Multidisciplinary perspectives on the origins and migrations of Austronesian-speaking peoples prior to 1000 BC. In Ben Roberts and Marc Vander Linden, eds *Investigating archaeological cultures: material culture, variability and transmission*. Dordrecht: Springer.

Forthcoming. Clues to the linguistic situation in Near Oceania before agriculture. In Tom Güldemann, Patrick McConvell and Richard A. Rhodes, eds *The languages of hunter-gatherers: global and historical perspectives*. Cambridge: Cambridge University Press.

2 Language in the present

2.1 Austronesian languages

1981 *The dialects of Sinagoro and Keapara. Unpublished manuscript. Canberra: The Australian National University.

1982 Aspect-marking in New Ireland: towards a historical reconstruction. In Rainer Carle, Martine Heinschke, Peter W. Pink, Christel Rost and Karen Stadtlander, eds *Gava': studies in Austronesian languages and cultures dedicated to Hans Kähler*, 173–196. Berlin: Dietrich Reimer.

1984 Maisin: a preliminary sketch. *Papers in New Guinea linguistics 23*, 1–82. Canberra: Pacific Linguistics.

1994 Describing inter-clausal relations in Takia. In Ger P. Reesink, ed. *Topics in descriptive Austronesian linguistics*. *Semaian 11*, 40–85. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.

1995 [Barbara F. Grimes, Joseph E. Grimes, Charles E. Grimes, Malcolm Ross and Darrell T. Tryon] Listing of Austronesian languages. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary,* Part 1, 121–279. Berlin: Mouton de Gruyter.

1995 Takia. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary,* Part 1, 677–685. Berlin: Mouton de Gruyter.

1995 Yabem. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary,* Part 1, 699–718. Berlin: Mouton de Gruyter.

1995 [Craig Throop and Malcolm Ross] Kaulong. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary.* Part 1, 719–726. Berlin: Mouton de Gruyter.

2002 [adapted by Malcolm Ross]. Kele. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 123–147. Richmond, Surrey: Curzon.

2002 Mussau. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 148–166. Richmond, Surrey: Curzon.

2002 [Joyce Sterner and Malcolm Ross]. Sobei. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 167–185. Richmond, Surrey: Curzon.

2002 [adapted by Malcolm Ross]. Kairiru. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 204–215. Richmond, Surrey: Curzon.

2002 Takia. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 216–248. Richmond, Surrey: Curzon.

2002 [adapted by Malcolm Ross]. Jabêm. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 270–296. Richmond, Surrey: Curzon.

2002 [Mike Anderson and Malcolm Ross]. Sudest. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 322–346. Richmond, Surrey: Curzon.

2002 [adapted by Malcolm Ross]. ‘Ala’ala. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 347–361. Richmond, Surrey: Curzon.

2002 Bali-Vitu. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 362–386. Richmond, Surrey: Curzon.

2002 [adapted by Malcolm Ross]. Kaulong. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 387–409. Richmond, Surrey: Curzon.

2002 Siar. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 410–425. Richmond, Surrey: Curzon.

2002 Taiof. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 426–439. Richmond, Surrey: Curzon.

2002 [adapted by John Lynch and Malcolm Ross]. Banoni. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 440–455. Richmond, Surrey: Curzon.

2002 Sisiqa. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages,* 456–466. Richmond, Surrey: Curzon.

2004 Aspects of deixis in Takia. In Gunter Senft, ed. *Deixis and demonstratives in Oceanic languages,* 15–36. Canberra: Pacific Linguistics.

2005 [Malcolm Ross and Stacy Fang-ching Teng] Clause constructions in Nanwang Puyuma. *Concentric: Studies in Linguistics* 31.119–158.

2005 [Malcolm Ross and Stacy Fang-ching Teng] Formosan languages and linguistic typology. *Language and Linguistics* 6.739–781.

2006 Formosan languages. In Keith Brown, ed. *Encyclopedia of language and linguistics.* 2nd edition, vol. 4, 587–590. Oxford: Elsevier.

2006 Malayo-Polynesian languages. In Keith Brown, ed. *Encyclopedia of language and linguistics.* 2nd edition, vol. 7, 457–462. Oxford: Elsevier.

2007 Two kinds of locative construction in Oceanic languages: a robust distinction. In Jeff Siegel, John Lynch and Diana Eades, eds *Language description, history and development: linguistic indulgence in memory of Terry Crowley,* 281–295. Amsterdam: John Benjamins.

2008 Negative verbal clause constructions in Puyuma: exploring constructional disharmony. In Alexander Bergs and Gabriele Diewald, eds *Constructions and language change,* 171–193. Berlin: Mouton de Gruyter.

Forthcoming. Loanwords in Takia, an Oceanic language of Papua New Guinea. In Martin Haspelmath and Uri Tadmor, eds *Loanwords in the world's languages: a comparative handbook*. Berlin: Mouton de Gruyter.

Forthcoming. Takia vocabulary. In Martin Haspelmath and Uri Tadmor, eds World Loanword Database. Munich: Max Planck Digital Library, 1218 word entries. [Available online: http://www.loanwords.info/Takia]

2.2 Papuan languages

1978 [Malcolm Ross with John Natu Paol]. *A Waskia grammar sketch and vocabulary*. Canberra: Pacific Linguistics.

1980 Some elements of Vanimo, a New Guinea tone language. *Papers in New Guinea linguistics 20*, 77–109. Canberra: Pacific Linguistics.

In press A preliminary analysis of East Kewa tone. In Kenneth McElhanon and Ger Reesinck, eds *A mosaic of language and culture: celebrations of the academic career of Karl Franklin*. Summer Institute of Linguistics.

2.3 Contact languages

1984 Current use and expansion of Tok Pisin: effects of Tok Pisin on some vernacular languages. In S.A.Wurm and Peter Mühlhäusler, eds *Handbook of Tok Pisin (New Guinea Pidgin),* 539–556. Canberra: Pacific Linguistics.

1992 [Malcolm Ross and T.E. Dutton] A note on Hees' 'Tolai-Nakanai' trade language. *Language and Linguistics in Melanesia* 23.198–204.

1996 Mission and church languages in Papua New Guinea. In S.A. Wurm, P. Mühlhäusler and D.T. Tryon, eds *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas.* 3 vols. Map 60 and 2.1:595–617 *Trends in linguistics, documentation* 13. Berlin: Mouton de Gruyter.

2.4 Cross-linguistic typology

1995 Diachronic typology and changing paradigms in historical linguistics: a review article of Johanna Nichols' *Linguistic diversity in space and time. Language and linguistics in Melanesia* 26.173–193.

2003 Typology and language families: a comment on Klamer's 'Typical features of Austronesian languages in Central/Eastern Indonesia'. *Oceanic Linguistics* 42.506–510.

2004 The morphosyntactic typology of Oceanic languages. *Language and Linguistics* 5.491–541.

2004 The grammaticization of directional verbs in Oceanic languages. In Isabelle Bril and Françoise Ozanne-Rivierre, eds *Complex predicates in Oceanic languages: studies in the dynamics of binding and boundness,* 297–330. Berlin: Mouton de Gruyter.

2004 Typologie morpho-syntaxique des langues océaniennes. In Elizabeth Zeitoun, ed. *Les langues austronésiennes,* 71–86. Paris: Ophrys.

2.5 General

1981 *Morpheme substitution and language function. Unpublished manuscript. Canberra: The Australian National University.

1994 [Malcolm D. Ross and Ingrid B. Ross] As clever as ever? *The Canberra Linguist* 24(2).14–19.

1995 [S.A. Wurm, T.E. Dutton, S. Holzknecht and M. Ross] Namen in Neu-Guinea. In *Namensforschung—Proper Name Studies—Les noms propres*. (Handbücher zur Sprachund Kommunikationswissenschaft), vol.1, 935–938. Berlin: Mouton de Gruyter.

3 Language teaching and education

1975 Skul bilong kirapim ples. *Yagl-Ambu* 2.148–157.

1975 Teaching the present perfect. *English in Papua New Guinea* 14.47–53.

1978 *A decentralised university? Paper presented to the Waigani Seminar, Port Moresby.

1978 *The analysis of interactive discourse and its relation to functional language teaching. Paper presented to the Twelfth Annual Congress of the Linguistic Society of Papua New Guinea.

1979 Using a reading passage to present grammatical structures, *Regional English Language Centre Journal,* Supplement 3.

1992 Doing linguistics with Icon. *NOAM: Notes on Apple Macintosh* 2(2).16–19, 22.

4 Edited volumes

1992 [ed.] *Papers in Austronesian linguistics No. 2.* Canberra: Pacific Linguistics.

1993 [Tom Dutton, Malcolm Ross and Darrell Tryon, eds] *The language game: papers in memory of Donald C. Laycock.* Canberra: Pacific Linguistics.

1993 [Malcolm Ross and Janet Ezard, eds] *Topics in the description of Kiriwina* by Ralph Lawton. Canberra: Pacific Linguistics.

1994 [Andrew Pawley and Malcolm Ross, eds] *Austronesian terminologies: continuity and change.* Canberra: Pacific Linguistics.

1996 [ed.] *Studies in languages of New Britain and New Ireland. 1*: *Austronesian languages of the North New Guinea Cluster in northwestern New Britain.* Canberra: Pacific Linguistics.

1996 [Mark Durie and Malcolm Ross, eds] *The Comparative Method reviewed: regularity and irregularity in language change*. New York: Oxford University Press.

1998 [Malcolm Ross, Andrew Pawley and Meredith Osmond, eds] *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 1: Material culture*. Canberra: Pacific Linguistics.

2001 [Andrew Pawley, Malcolm Ross and Darrell Tryon, eds] *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Canberra: Pacific Linguistics.

2002 [John Lynch, Malcolm Ross and Terry Crowley, eds] *The Oceanic languages*. Richmond, Surrey: Curzon Press.

2002 [Fay Wouk and Malcolm Ross, eds] *The historical and typological development of western Austronesian voice systems.* Canberra: Pacific Linguistics.

2003 [Malcolm Ross, Andrew Pawley and Meredith Osmond, eds] *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 2: The physical world.* Canberra: Pacific Linguistics.

2005 [Wayan Arka and Malcolm Ross, eds] *The many faces of Austronesian voice systems: some new empirical studies.* Canberra: Pacific Linguistics.

2008 [Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross, Ilia Peiros and Marie Lin, eds] *Past human migrations in East Asia: matching archaeology, linguistics and genetics*. London: Routledge.

2008 [Malcolm Ross, Andrew Pawley and Meredith Osmond, eds*] The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 3: Plants.* Canberra: Pacific Linguistics.

5 Reviews

1976 Review: *Education in Papua New Guinea,* by Geoffrey Smith. *Comparative Education Review* 20.252–255.

1978 Review article: *The Tigak language of New Ireland*, by C.H. Beaumont. *Kivung* 11.184–192.

1985 Review: *Holle lists: vocabularies in languages of Indonesia*, vols 1 and 2, edited by W.A.L. Stokhof. *Language and Linguistics in Melanesia* 13.106–110.

1987 Review: *A grammar of Manam*, by F. Lichtenberk. *Journal of the Polynesian Society* 96.134–135.

1987 Review: *Kilivila: the language of the Trobriand Islanders*, by G. Senft. *Canberra Anthropology* 10.86–90

1988 Review: *The history of the Fijian languages*, by Paul Geraghty. *Language and Linguistics in Melanesia* 17.161–165.

1989 Review: *Studies in Austronesian linguistics*, edited by R. McGinn. *Asian Studies Association Review* 13.

1992 Review: *Pacific Island languages: essays in honour of G.B. Milner*, edited by J.H.C.S. Davidson. *Language and Linguistics in Melanesia* 23.68–73.

1995 Review: *Sprachkontakt und Sprachwandel im Baskenland*, by Martin Haase. *Linguistics* 33.618–624.

1995 Review: *The formation of Afrikaans*, by Paul T. Roberge. *Notes on Linguistics* 70.52–55.

1996 Review: *Germano-European: breaking the sound law*, by Tony D. Griffen. *Notes on Linguistics* 73.42–49.

1996 Review: *The Yoruba koiné – its history and linguistic innovations,* by J. Gbenga Fagborun. *Notes on Linguistics* 75.47–51.

1998 Review: *Language change in child and adult Hebrew* by Dorit Diskin Ravid. *Notes on Linguistics* 80.52–55.

1998 Review: *The Nalik language of New Ireland, Papua New Guinea*, by Craig Volker. *Oceanic Linguistics* 37.366–374.

1999 Review: *Linguistic reconstruction and typology,* edited by Jacek Fisiak, and *Linguistic typology, universality and the realism of reconstruction,* by Frederick W. Schwink. *Linguistic Typology* 3.134–144.

2000 Review: *Selected papers from the Eighth International Conference on Austronesian Linguistics*, edited by Elizabeth Zeitoun and Paul Jen-kuei Li. *Oceanic Linguistics* 39.445–455.

2001 Review: *Diachronic syntax: models and mechanisms*, edited by Susan Pintzuk, George Tsoulas and Anthony Warner. *Linguist List* 12.1886, 24th July.

2003 [John Bowden and Malcolm Ross] Review essay: Trilingual dictionary of Lamaholot (Indonesia) [Review: *Koda kiwã: dreisprachliches Wörterbuch des Lamaholot (Dialekt von Lewolema)*, by Karl-Heinz Pampus. Stuttgart: Deutsche Morgenländische Gesellschaft/Franz Steiner Verlag, 1999.] *Asia–Pacific Journal of Anthropology* 4.143–149.

2003 Review: *Serial verbs in Oceanic: a descriptive typology,* by Terry Crowley. *Journal of the Polynesian Society* 112.409–413.

2004 Book notice: *Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia*, by N.J. Enfield. *Language* 80.884–885.

2004 Book notice: *Paiwan-Wörterbuch: Paiwan–Deutsch, Deutsch–Paiwan*, by Hans Egli. *Language* 80.884.

2005 Book notice: *Javanese grammar for students (revised edition)*. By Stuart Robson. *Language* 81.1019.

2005 Review: *Otto Dempwolff's grammar of the Jabêm language in New Guinea*, translated by Joel Bradshaw and Francisc Czobor. *Anthropological Linguistics* 47.447–450.

2006 Review: *Tobelo*. by Gary Holton (Lincom Europe). *Linguist List* 17.292, 27th January.

2006 Review article: *Examining the farming/language dispersal hypothesis*, edited by Peter Bellwood and Colin Renfrew. *Language* 82.628–648.

2007 Book notice: *A formal–functional analysis of language, with an application to Ibanag*, by Eugene. *Language* 83.475.

2009 Review: *10-ICAL historical-comparative papers*, edited by David Mead. *Oceanic Linguistics* 48.274–286.

2009 Review: *Language and text in the Austronesian world: studies in honour of Ülo Sirk,* edited by Yury A. Lander and Alexander K. Ogloblin. *Linguist List.* 20.1943, 20th May.

Forthcoming book notice: *Language contacts in prehistory: studies in stratigraphy*, edited by Henning Andersen. *Language.*

Forthcoming book notice: *Germanic tone accents: proceedings of the First International Workshop on Franconian Tone Accents, Leiden, 13–14 June 2003*, edited by Michiel de Vaan. *Language.*

Forthcoming book notice: English intonation: an introduction. By J.C. Wells. *Language.*

Forthcoming book notice: Language choice and identity politics in Taiwan. By Jennifer M. Wei. *Language*.

6 Dissertations

1963 Medieval political and satirical verse. B.A. (Hons.) dissertation, Department of English, University of Bristol.

1972 German, Russian and communist elements in education in East Germany. M.Litt. (Education) dissertation, School of Education, University of Bristol.

1986 Proto Oceanic and the Austronesian languages of western Melanesia. PhD dissertation, Department of Linguistics, The Australian National University.

Part I

Historical relationships amongst languages

3 *Remapping the Austronesian expansion*

ROGER BLENCH

1 Are language phyla valid concepts beyond linguistics?

One of the curious by-products of historical linguistics is that its constructs tend to take on a life of their own. Something that begins life as a purely linguistic hypothesis comes to accrete other disciplines; reconstructions of the hypothetical Indo-European language lead to people called the Indo-Europeans, whose lifestyle can be imagined and whose genetics can be investigated. In some ways this seems reasonable; if there is manifest evidence for the spread of a language family then it should presumably reflect some sequence of events in the real world, perhaps a migration of peoples or of cultural ideas. If a demographic expansion is proposed, then in principle it should have both archaeological and genetic correlates, if these can only be identified. However, modern analogues can also point to the problems that arise from this way of thinking. English is an Indo-European language, and is now spoken in many geographically and culturally diverse regions, reflecting a variety of military and sociological imperatives. An external researcher with no knowledge of these background elements might draw highly inaccurate conclusions from modern-day contexts of English when reconstructing a hypothetical protoculture. It has been only partly humorously proposed that on purely synchronic evidence, Tok Pisin, rather than Hittite, would be the primary branching of Indo-European.

While mindful of these possibilities of error, it is also possible to be over-cautious. Languages disappear and are assimilated, and to bound the concept of a phylum by the geographical distribution of recently spoken languages would be to recover a very limited image of a broader historical process. In pre-European times, Pama-Nyungan languages covered nine-tenths of continental Australia, but given the known antiquity of the settlement of Australia, it would be perverse not to imagine they replaced or assimilated a greater diversity of pre-existing languages of unknown affiliation.

This paper examines the possibility that we should take a broader view of the expansion of the Austronesians. Studies in Austronesian linguistics are dominated by a characteristic map, looping from Madagascar to Easter Island, and reconstructions of Austronesian culture confine themselves to forms derivable from existing languages. This is perfectly acceptable *as a purely linguistic process*, but rather limited as a contribution to human history. A subset of linguists and archaeologists accept the Austronesian peoples as a historical reality, and assume they were highly mobile, making use of advanced maritime technology. It therefore seems reasonable to assume that they were able to reach many

places other than those where their languages are spoken today. And there appears to be some *a priori* evidence that this occurred. This paper[1] reviews the evidence for now-vanished Austronesian visits and settlement in a wide variety of locations across the world with a view to a conceptual remapping of the Austronesian expansion. Archaeology, transfer of crops and material culture and historical records can all contribute to redrawing the map of Austronesian migration. Some of these claims are more controversial than others, but the paper draws together and assesses claims for Austronesian presence in: (i) Japan; (ii) China; (iii) Thailand/Myanmar; (iv) Pacific Islands; (v) Australia; (vi) India and Sri Lanka; (vii) Indian Ocean islands; (viii) The Persian Gulf; (ix) Madagascar and the East African coast; (x) West Africa; (xi) South Africa; and (xii) the New World. Also included are some examples of later migrations, such as the Javanese in Surinam, where an Austronesian language is still extant but is disappearing, to illustrate the broader processes whereby demographic transfer can rapidly become dissociated from a current spoken language.

It should also be possible to distinguish chronological layers within this scenario, for example, those hypothesised in prehistory and those with fairly secure historical documentation. For example, contact between the Austronesians and the New World is inferred from indirect evidence and its dating is controversial and can only be assigned to very broad time periods. However, the presence of Austronesians in the Persian Gulf is recorded in Arabic chronicles which are securely dated and the years of these events can be assigned. The movement of Javanese to Surinam is extremely well-documented and members of these communities can still be interviewed today. This suggests that we can assign at least three types of chronological certainty to Austronesian expansions.

While links between Austronesian languages, such as the similarities between Malagasy and Malay, were noted as early as the seventeenth century, Austronesian undoubtedly begins with Dempwolff in the 1930s, although his *Austronesisch* is what today would be called Malayo-Polynesian or extra-Formosan. It was probably first picked up as an archaeological concept by Peter Bellwood in the 1970s (e.g. Bellwood 1978) and since then has developed rich associations in different disciplines. Books such as *The Austronesians* (Bellwood et al. 1995) assume the reality of their culture across a wide variety of disciplines. Which is not to say detractors do not exist; Solheim, Meacham and Terrell being major names (Solheim 1984–85; Meacham 1984–85, 1991, 1995, 2004; Terrell 2004; Terrell et al. 2001). A rather different challenge is presented by writers such as Oppenheimer (2004) who present one view of what is claimed to be the genetic evidence. This paper takes the view that these authors simply do not engage with the linguistic evidence sufficiently thoroughly and that their alternative models do not account effectively for the data.

---

[1] My interest in the Austronesians has been stimulated over the years through discussions with Malcolm Ross at a variety of conferences and the invitation to contribute to this Festschrift seemed a good opportunity to synthesise the scattered material I have been gathering over the years. I was able to present a very early version of this material to the Archaeology Department at the University of the Philippines in 2006 and I would like to thank Victor Paz for the invitation and subsequent feedback. I would also like to thank Sander Adelaar, Robert Blust, Mark Hudson, Malcolm Ross, Laurie Reid, Laurent Sagart, Glenn Summerhayes, Matthew Spriggs and Martin Walsh for sending me unpublished material contributing to my thinking on these issues. Thanks also to the anonymous referees for pointing me in the direction of some material I might otherwise have missed. I would especially like to emphasise that these individuals are in no way responsible for some of the hypotheses floated here, which intentionally speculate outside the mainstream.

A key issue in the Austronesian expansion debate is between the interdisciplinary archaeologists such as Peter Bellwood who broadly support a 'Neolithic' package which includes demic expansion, agriculture, pigs, chickens and certain types of artefact such as the *lingling-o*, and a more resistant archaeological coterie who claim the diversity of material culture on the ground does not support such a simple model (e.g. Szabo and O'Connor 2004; Anderson 2005). In some ways the linguists tend to side with the unadorned Austronesian expansion model (e.g. Pawley 2003; Pawley and Ross 1993; Gray et al. 2009) though sceptical voices are being raised (e.g. Donohue and Denham in press). Recent evidence from a so far unexploited source of evidence, bacteria varieties co-associated with distinct human populations has provided intriguing confirmation for both the Austronesian expansion, the clear distinction from Papuan and Australian populations and a link with the Chinese mainland (Moodley et al. 2009).

Austronesian languages are manifestly in geographically dispersed locales; the likelihood that this is just some sort of cultural diffusion seems improbable. After all, in recent history the spread of languages has certainly been riding a carrier wave of actual migration. On the other hand, the situation on the ground will always be more complex than such a model implies; migrants are likely to interact with a diversity of existing cultures and produce a wide variety of outcomes. In that sense, the burden of this paper may seem even more problematic to such archaeologists; for the evidence is undoubtedly highly fragmentary. Nonetheless, there is a clear value in compiling it, to open up the Austronesian debate.

2 The further adventures of the Austronesians

Two very fundamental questions can be asked of Austronesian culture; where did it originate and where did Austronesian navigators reach? Although it is broadly accepted that Taiwan is the 'homeland' of existing Austronesian languages, archaeological evidence suggests strongly they are an incoming population. So where was their ultimate homeland? The Austronesian expansion crested a wave of pioneer voyages, populating islands as far-flung as Madagascar and the Marianas. To assume their presence elsewhere would not be unreasonable. If Austronesian languages are no longer spoken in a particular location today, then what type of evidence might there be for their former presence? Table 1 lists some of the possible categories of evidence.

Table 1: Categories of evidence for Austronesian contact

Category	Type	Example
Linguistic	Loanwords	South America
	Historical testimony	East African coast, Arabia
Biological	Introduced plants and animals	Central America
	Zoogeography	Myanmar
	Bacteria geography	Island SE Asia
	Disease	Africa
Genetic	Distinctive lineages, iconography	South America
Archaeological	Material culture, settlement patterns	Uninhabited Pacific islands, Australia
Ethnographic	Material culture	Maldives, East Africa

The purpose of this paper is to draw attention to hypotheses concerning Austronesian contact with regions where Austronesian languages are no longer spoken. These include: (i) Japan; (ii) China; (iii) Thailand/Myanmar; (iv) Pacific Islands; (v) Australia; (vi) India and Sri Lanka; (vii) Indian Ocean islands; (viii) The Persian Gulf; (ix) Madagascar and the East African coast; (x) West Africa; (xi) South Africa; and (xii) the New World. Some of these claims are more controversial than others, and I expect at least some to be comprehensively falsified in the future. There is moreover, a not inconsiderable literature, especially on trans-Pacific contacts, which is close to the lunatic fringe. I can only hope to avoid a compilation such as this being summarily consigned to the outbox. The object is to help reconceptualise the Austronesian expansion as something more than a linguistic hypothesis and to suggest new avenues of research and comparison to specialists in particular geographic areas. It should not be assumed that all these possible movements were at the same period; absolute dating is the province of archaeology. For example, the movement to the Yaeyama islands, almost certainly from Taiwan (§3.2) seems to have been quite recent, despite the relatively short voyage necessary. By contrast, archaeology suggests a much older voyage to the remote Marianas (§4.2), indicating a different phase in the technological evolution of Austronesian culture.

3 Japan

3.1 Austronesians on the Japanese islands

The evident capacity of the Austronesians to reach far-flung places has long encouraged scholars to propose that they could also have reached Japan. A chain of small islands joins the northeast corner of Taiwan to the main Japanese islands, so navigators on Taiwan might also have sailed north while sailing east and south. As a consequence, there is quite a venerable literature imagining an Austronesian origin for Japanese. Although the idea was propounded as early as 1911, its most eloquent exponent has been Ōno Susumu (1970) who believed that Japanese was an Austronesian language with a later Altaic superstrate. He later added Tamil to this unfortunate roll-call, by which time the move into science fiction was almost complete. Hudson (1999) provides a concise account of the evolution of these ideas together with some indication of the ethnic and nationalist agendas that may lie behind them.

However, such ideas were not confined to Japanese scholars. Paul Benedict, much revered scholar of Sino-Tibetan, published a late book, ‘Japanese-Austro-Thai’ (Benedict 1990) in which he gave lexical evidence for linking Austronesian, Daic and Japanese. Despite the apparent wealth of examples, each proposed cognate depends on significant special pleading, always the sign of a doubtful ‘long-range’ hypothesis. Needless to say, the mainstream Austronesian establishment has not followed Benedict’s lead and Vovin (1994) published a comprehensive demolition of Benedict’s arguments. Broadly speaking, neither linguistics nor archaeology have provided any support for the notion that Austronesian mariners reached Japan proper.

A rather different approach to this issue is found in the writings of Ann Kumar (1998, 2007) and Kumar and Rose (2000). Kumar is convinced that there was early contact between Java and Japan and this idea is buttressed with evidence from rice genetics (Morinaga 1968), culture words and similarities of notions of kingship. The linguistic evidence is a series of comparisons between Old Javanese and Old Japanese, some more

credible than others, of words from marginal vocabulary, often reduplicated (Kumar and Rose 2000). While it is not impossible that there was some contact between Japan and Java in the early medieval period, the type of pervasive and early influence posited here has not converted other researchers.

The Austronesians have such a distinctive material culture signature that early incursions should surely be evident. There is, however, one striking exception, the Yaeyama islands (八重山諸島 *Yaeyama-shotō*), the southernmost in the Japanese chain politically, but geographically much closer to Taiwan.

3.2 The Yaeyama islands

The Yaeyama islands lie some 100 km east of northern Taiwan and it seems positively unlikely that they were not reached by Austronesians in view of the documented expansion towards the Marianas. However, they have been little known archaeologically until recently, partly because they lie outside the Jōmon area, which is typically the focus of mainstream Japanese archaeologists. Summerhayes and Anderson (in press) summarise Japanese publications on Shimotabaru wares, which are possible evidence for Austronesian colonisation from Taiwan, 4000–3800 BP. Comparisons with Taiwanese pottery and axe forms taken to be a signature of Austronesian expansion suggest striking parallels, although Shimotabaru wares are much reduced in decorative terms. There is, moreover, no unambiguous evidence for agriculture, although it seems likely that pigs (*Sus scrofa riukiuanus*) were translocated at this period. Further north, Hudson et al. (submitted) have investigated the site of Nagabaka on Miyako island, and recorded four radiocarbon dates from the bottom level of the midden which range from 1520 to 1215 cal BP and were associated with two *Tridacna* adzes. Such a date is disappointingly recent compared with the dates from the southern Yaeyamas (Iriomote) and there is furthermore no sign of agriculture. As Hudson et al. suggest, these sites could be the result of an accidental drift voyage and temporary foraging settlement, eventually abandoned.

3.3 Ships that pass in the night

None of the literature so far seems to have followed what might seem an obvious path, namely to compare the Japanese dialects spoken in the far south of the island chain with Taiwanese Austronesian for possible substrate influence or even loanwords. Comparisons are entirely with putative Old Japanese or variously reconstructed versions of mainstream dialects. The language of Yonaguni island, for example, appears to conserve phonological features of archaic Japanese. Nonetheless, it is quite probable that even if Ryukyuan were investigated, it might not produce clear evidence for the intersection of cultures as the Austronesians may have deserted the islands before the expanding proto-Japanese encountered them.

4 Austronesians in China

4.1 China as a source for the Austronesians

The Austronesians are generally considered to have reached Taiwan by ca. 6000 BP, gradually eliminating or assimilating the Pleistocene populations (Tsang 1995, 2001; Rolett et al. 2002; Rolett 2007; Bellwood 2007; Blench 2008a) and leading to the highly

distinctive pattern of languages found there today. Almost all scholars agree on the Chinese mainland as their source region and there has been a wide body of support for a site such as Hemudu (河姆渡文化), a Neolithic culture that flourished just south of the Hangzhou Bay in Jiangnan in modern Yuyao, Zhejiang. There seem to be some problems with this view (e.g. Anderson 2000). Laurent Sagart (pers.comm.) has argued that various aspects of Austronesian culture on Taiwan point to a source region significantly further north. One reason for this is that it is millet, not rice, that lies at the heart of Formosan agricultural rites, whereas the Hemudu horizon is distinctively rice-based (though see Fuller et al. 2008 who argue that the rice at Hemudu may not have been domestic). Whatever the case, it points to a significant original Austronesian population on the Chinese mainland, which presumably accounts for a wide range of cultural similarities with Austroasiatic and other minority ethnolinguistic groups. The Austric hypothesis, whereby Austronesian and Austroasiatic constitute a macrophylum has been gaining some support in recent years. Blust (1996) for example, has argued that Austric must have its *Heimat* in Leaping Tiger Gorge, Yunnan. Even supporters of Austric have yet to follow this particular bound. However, no trace of these Austronesian populations appears to remain linguistically; no substrate in Chinese dialects has ever been identified. A macrophylic view strongly developed by Sagart (1994, 2005a) is that Sino-Tibetan and Austronesian are genetically related and cognate items occur in fundamental vocabulary, although more with Sinitic than Sino-Tibetan as a whole.

4.2 The Daic hypothesis

The Daic or Tai-Kadai languages cover a substantial region of East and SE Asia. Thai, their best-known representative, dominates Thailand, but the Daic languages are generally considered to originate in South China, where they are most diverse (Edmondson and Solnit 1988). Despite their importance, little is known about their prehistory, homeland and the causes of their expansion; proposed archaeological correlations deal only with the most recent phases.

All the language phyla of East Asia have been argued as connected with one another at different times. Early ‘Indo-Chinese’ hypotheses linked Daic with Chinese, or later, Sino-Tibetan (van Driem 2005). Influential for a long period was ‘Austro-Thai’ first advanced by Benedict (1942, 1975), which broadly claimed Austronesian and Daic were related. Benedict (1990) later expanded his view to include Japanese, a direction in which few have followed. A problem for many authors was that Daic and Austronesian surface morphologies appear to be very different; Daic is highly tonal with very short words, Austronesian is non-tonal and tends to have CVCV stems plus affixes. Hence the tendency was to treat it as isolated or to link it with Sino-Tibetan, which appears much more similar in terms of morphology. Thurgood (1994) argues that the relation with Austronesian is simply that of loanwords.

Benedict is often criticised for irregular semantics and individual arguments for each form, which lowers the threshold for a demonstration of relatedness. However, Ostapirat (2005) makes a more convincing argument for a genetic relationship between Daic and Austronesian based on regular sound-correspondences. Ostapirat does not advance a hypothesis as to the place of Daic within Austronesian, as his paper links ‘proto-Kra-Dai’ with the Austronesian reconstructions of Dempwolff and Blust. Sagart (2004, 2005b), following this line of argument, places Daic on a level corresponding to Malayo-Polynesian as a branch of ‘Muish’, part of his proposed phylogeny of Formosan Austronesian. Indeed

the evidence Sagart cites from Buyang, a mainland Daic language, shows that typical Austronesian morphology was conserved *after* the arrival of speakers back on the mainland and that the reduced forms now typical of most Daic languages are a later development.

If this linguistic scenario is accepted, then proto Daic speakers would have migrated back from the southern Taiwan about 4000 BP, at the same time that other Austronesian speakers were colonising the northern Philippines and only slightly earlier than they reached the Marianas, apparently aided by newly developed maritime techniques (Hunter-Anderson et al. 1995). At a period of significant maritime dispersal, there is no reason in principle against such a back-migration. All the diversity of Daic languages is in China: despite the southward extension of Thai today, the likely origin of Daic is in Guizhou, although Daic languages were presumably formerly spoken in Guangdong and have now been assimilated by Sinitic. One possible confirmation of this is a cluster of features in material culture and iconography between the cultures of aboriginal Taiwan and the Daic peoples, such as the blackening of teeth, dental ablation, multi-tongue jew's harps and the motif of intertwined snakes (Blench 2008c).

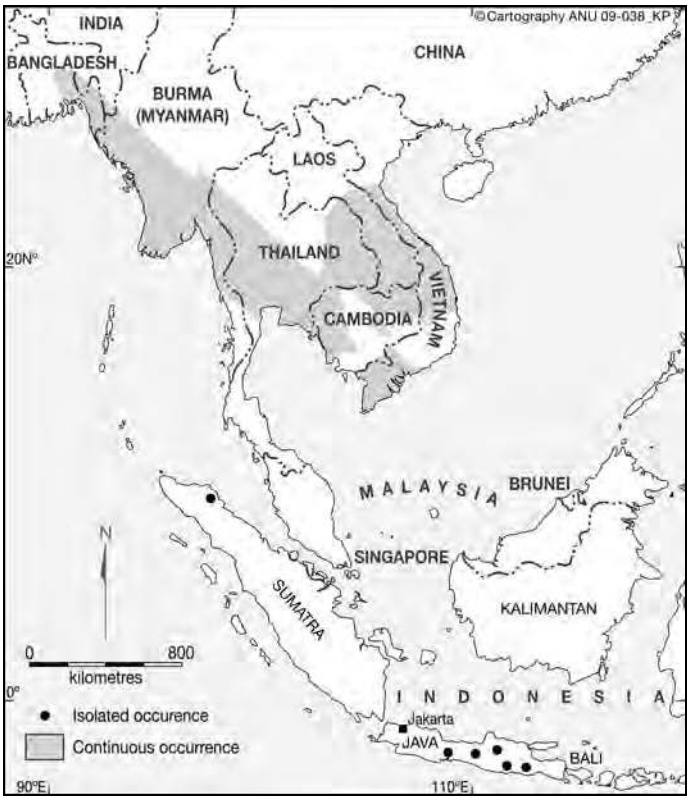
5 Myanmar/Thailand

The northernmost Austronesian presence in the Gulf of Thailand today is the Moken boat people who live primarily on boats in the Mergui Archipelago, Dung, and other islands in south Myanmar. There are currently some 7000 Moken in Myanmar. Moken is a Malayic language most closely related to Moklen, spoken further south in the same area. However, given the rich resources of the Andaman Sea and the evidence for Austronesian voyages towards India, it would not be surprising if outriggers ventured further north into the Gulf of Thailand.

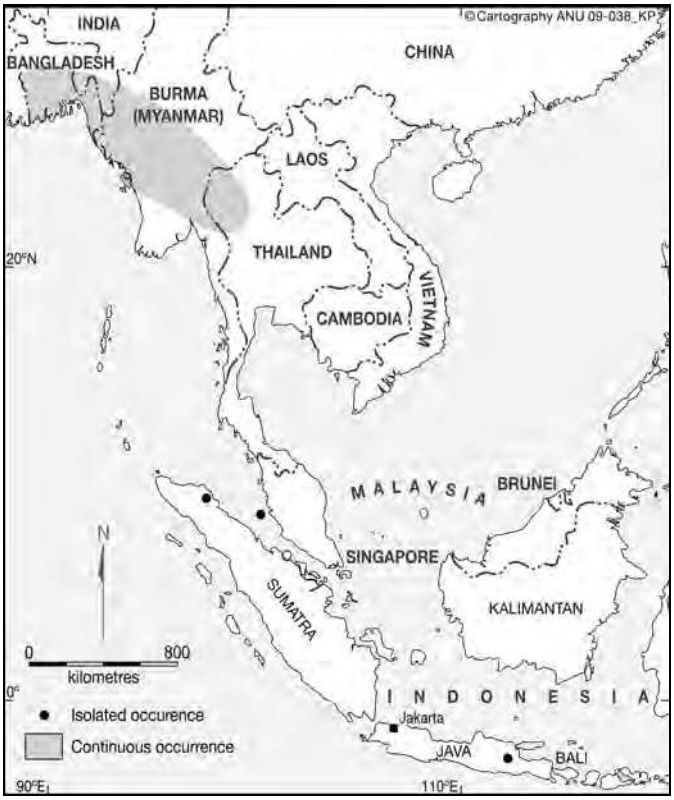
A distinctive trail of evidence for such contact is provided by the zoogeography of commensals and domestic animals. Groves (1995) studied the distribution of ricefield commensals such as *Mus cervicolor* and the bandicoot-rat, *Bandicota bengalensis* (see Maps 1 and 2), both of which have intriguing distributions. The mainland distributions are in Myanmar and Thailand and appear to reflect intensive rice-cultivation. Both species are commensals strongly associated with agriculture. The individual records further south point to sea-borne translocation, presumably unintentional and point to Austronesian trading voyages along the Burmese/Thai coast at an unknown point in the past.

6 Pacific islands

Perhaps the most obvious case for evidence of past Austronesian presence is on now-uninhabited Pacific islands. Anderson (2002), in a survey of Pacific islands in remote Oceania, found no less than thirty apparently reached by Austronesian navigators but which were devoid of inhabitants when first reached by European explorers. These were approximately: 5 in the Pitcairn-Henderson Islands region; 8 in the New Zealand region; 1 in New Caledonia; 1 in Tonga; 3 in the Cooks-Societies; 8–9 in the Equatorial islands region; and 4 in Hawai'i. The material evidence consists of platforms (*ahu*), shell axes and fish-hooks and other culture items typical of Austronesians. However, there is also substantial evidence for landscape modification and faunal collapse on islands with no archaeological sites. Various explanations have been advanced for their desertion; typhoons, disease, lack of sustainable food and water supplies are all possible causes.



Map 1: Distribution of *Mus cervicolor*



Map 2: Distribution of *Bandicota bengalensis*

7 Australia

Australia is a large stretch of land, difficult to miss for maritime peoples, and its northern coasts are not unwelcoming for tropical agriculture. It seems hard to imagine that the expanding Austronesian speakers did not encounter it at some point in their journeys. Late (i.e. 13th century) dates for the settlement of New Zealand make it more explicable that Polynesians did not reach the Eastern seaboard.

Malay trepangers were visiting the northern coast of Australia from the 17th century onwards in search of sea-cucumbers (Macknight 1976; Ganter et al. 2006). But there have been claims that linguistic evidence shows Austronesian voyagers must have been landing on the coast of Northern Australia far earlier, as evidenced by deep-level loanwords in Pama-Nyungan languages (e.g. O'Grady and Tryon 1990). Plant material may also reflect this earlier Austronesian presence. Feral taros have been recorded in Arnhem Land which may have been left by these earlier voyagers (Levitt 1981), be an earlier Papuan introduction or be part of the indigenous flora. Denham, Donohue and Booth (2009) propose that it represents a record of indigenous horticultural experimentation. An endemic bamboo, *Bambusa arnhemica*, found across parts of Northern Australia, has no mainland relatives and shows links to Asiatic bamboos (Franklin 2003). Two other possible pieces of evidence may be relevant; the introduction of the dingo and the use of outriggers in Cape York. The dingo, a subspecies of an East Asian dog, ultimately descended from the wolf, seems to have arrived in Australia 3500–4000 BP (Koler-Matznick 2002). Its presumed source was Timor, although there is no direct evidence for this. The dates also seem slightly early for an Austronesian maritime source, but surprisingly late for Papuan contact. Early surveying voyages, such as HMS Rattlesnake in 1848, recorded sophisticated outriggers in the Cape York Peninsula, resembling those of the Torres Straits islands (Flood 2006). The presumption is that these must have ultimately had an Austronesian origin, although possibly via Papuan intermediaries, rather than directly.

Even better would be evidence such as identifiable pottery sherds. Although rumours have surfaced of Lapita fragments found in Australia, none have ever been formally published. However, it is the case that some offshore islands, such as Keppel Island, have unusual archaeological cultures. Rowland (1981, 1982, 1984, 1987) who excavated these islands, never posits Austronesian contact, but his studies on the idiosyncratic fish-hooks and water-craft certainly point to 'culture contact'. It seems at least possible the east coast of Australia was the end-point of experimental Austronesian voyages but for some reason, perhaps climatic, the travellers never settled permanently.[2]

8 India and Sri Lanka

India is relatively close to the Austronesian world and there is abundant historical evidence for the sea movement of imperial India across the Andaman Sea to SE Asia, with the formation of the so-called 'Indianised' States from at least the 6th century onwards (Ray 1989, 2003 and Munoz 2006). However, it would seem unlikely that there was no traffic in the opposite direction. There is no direct archaeological or linguistic evidence for this, but as so often, it is not clear in whose interest it would be to look for such evidence. Indirect evidence can be subject to conflicting interpretations; shared cultural traits can result from

[2] Oscar Wilde on America; 'Of course America had been discovered many times before Columbus, but it had always been covered up'.

single introductions or be brought by intermediaries such as traders. However, there would seem to be abundant evidence for shared cultural traits. Hornell (1920) noted striking agreements in the construction of boat-types to suggest 'Polynesian' influence in India, and also noted the cultural context of coconut cultivation and toddy tapping. Waruno Mahdi (1999a, b) has synthesised textual references and evidence for shipping types. The argument is long and intricate, but the conclusion is that the Nāgas referred to in early texts 'typically inhabited islands, the sea coast or banks of rivers. Some of them worshipped megaliths and practised buffalo sacrifice and head-hunting' (Mahdi 1999b:182). Identifying such populations with early Austronesian migrants would not be unreasonable. Given the dates for the texts, Austronesian presence would be identifiable from ca. 500 BC.

There is an old argument for the spread of the clove from insular SE Asia to India. Sanskrit *lavaŋga* (लवङ्ग) is claimed to be cognate with Old Javanese *buŋa-lawaŋ* (Donkin 2003) and Mahdi (1999b) also proposes other relationships including words for 'lime' and 'camphor'. However, there is also phytogeographical evidence for other fruit species (Blench 2008b), for example the bilimbi and carambola (*Averrhoa spp.*), the lime (*Citrus aurantifolia*), the coconut (*Cocos nucifera*), the langsat (*Lansium domesticum*), the noni (*Morinda citrifolia*) and the santol (*Sandoricum koetjape*). Another fragment of related evidence comes from the distribution of the blow-gun (§15). Although it clearly developed in the Austronesian world, it has a relic distribution in South India and Sri Lanka. Indirect pointers to Austronesian settlement in SE India are quite strong, but archaeological and linguistic evidence remains weak. It seems likely that numbers were never high[3] and journeys to India represented opportunistic trade and piracy rather than a concerted attempt to settle, hence the somewhat difficult to interpret textual evidence.

In contrast, much more recent arrivals are represented by the Orang Melayu of Sri Lanka (Adelaar 1991). This Muslim community of about 50,000 persons is descended mainly from Javanese political exiles, soldiers and convicts, who came from Java during Dutch colonial rule, 1658–1796 (Hussainmiya 1987). Although most Sri Lankan Malays are of Javanese ancestry, they also originate on other Indonesian islands such as Bali, Tidor, Madura, Banda and Ambon.

9 Indian Ocean islands

If indeed there were early and persistent Austronesian contacts across the Indian Ocean, both via outriggers as part of the spice trade and with the rigid ships of the Malay sea-borne empire, it seems curious there is no evidence for settlement of islands in the middle of the ocean (Blench 2007, in press). The ability of Austronesian navigators, who would have been responsible for the spice trade to the Roman Empire well before the rise of Malay shipping, to find very small islands in large expanses of open ocean is well documented, yet it seems that almost all the Indian Ocean islands were uninhabited at first European contact. Archaeology has so far produced no evidence for Austronesian (or other) incursions on islands such as Diego Garcia, the Seychelles and Mauritius. The situation in the Pacific may have been replicated in the Indian Ocean; Mauritius or the Seychelles may have been reached, but then abandoned. One reason may have been lack of easily exploitable food resources; European sailors tended to rapidly consume any readily caught resource (e.g. the dodo) and bring in and release often destructive species such as

[3] Although Mahdi (1999b:168) quotes the *Kiṣkindhākāṇḍa* as saying the migrants 'live on the sea's milky beach, and in the *tamāla* woods live, and of coconuts eat, their number is countless'.

the goat in order to ensure future food supplies. Similar depredations leading to collapse of bird faunas in the Pacific, most notably the moa in New Zealand, have been well-documented (Anderson 2002).

The Maldive archipelago some 600 km south-east of the Indian mainland is today inhabited entirely by speakers of Divehi, an Indo-Aryan language. Although the dominant religion is presently Islam, there is strong evidence for numerous prior visitors from other cultural spheres, most notably Buddhism (Heyerdahl 1986). Archaeology in the Maldives is still underdeveloped, and so far no sites point to Austronesian presence. However, Hornell (1920:230) observes that constructional techniques in boatbuilding point unambiguously to early Austronesian ['Indonesian' in his terms] contact. Manguin (1993:265) notes 'field work in the Maldives, … found evidence to prove that the shipbuilding tradition there used to be of the Southeast Asian sewn-plank and lashed-lug type (as opposed to an Indian Ocean sewn-plank tradition'.[4] It seems increasingly likely that the absence of evidence for Austronesian landings is an artefact of the exiguous archaeology.

10 The Persian Gulf

Although Austronesian navigators may also have reached the Persian and Arabian Gulfs as part of their expeditions west across the Indian Ocean, there is no immediate archaeological evidence for this. But textual evidence from the post-Islamic period provides a more explicit record of their presence. Goeje (1894) was the first to identify reports of the *Sayābiğa* (Sumatrans) settling in the Persian Gulf and Ferrand (1934) expanded these references. Balâduri (9th century) and al-Tabarî report that Sayābiğa were in Sind prior to the expansion of Islam in this region.[5] Isolated from the connection with their home area, they became mercenaries. During the reign of the Calif Abu Bakr (632–634) the Sayābiğa garrisoned Bahrein, while in 656–657 AD, they were employed to guard the treasury at Basra. In 775, Sayābiğa are recorded as taking part in a naval expedition against the coast of NW India. It is this type of flow among mariners that must have been responsible for the interchange of maritime terminology between Malay, Arabic and Swahili (Blench in press).

11 Madagascar and East Africa

Austronesian presence in Madagascar is uncontroversial, since the Malagasy language is Austronesian and closely related to Barito languages of Borneo. Nonetheless, its exact genealogy has been subject to considerable debate (Beaujard 2003). Many areas of vocabulary seem to be borrowed from Malay, in particular, sailing terminology (Adelaar 1996). Blust (2005) has shown that the languages of the Samal or Bajau Laut, the nomads still traversing the seas between NE Borneo and the southern Philippines, are part of the Barito group. The earliest evidence for Austronesian settlement of Madagascar is the 5th century AD, although this date is uncertain and only by the 7th century is evidence uncontroversial (Blench 2007). It seems possible that this reflects the expansion of the Srivijaya Malay in the 7th century.

---

4 It is worth emphasising there are distinctive boat types and shipbuilding techniques in the Maldives not attested on the Indian mainland.

5 I am indebted to Philippe Beaujard for this material which will be incorporated in a forthcoming major study he is preparing.

However, if ships could reach Madagascar they could and indeed should have also reached the East African coast, although no Austronesian language is spoken there today. Nonetheless, there is significant cultural evidence for Austronesian presence on the East African coast, some of it well before the settlement of Madagascar (Adelaar 2006, in press; Blench 1996, 2009b and in press). Pliny (ca. 1 AD) refers to the 'men who come across the great ocean on rafts [*rati*]' which he contrasts with coastal traders (Rackham 1942). He describes them as traders in spices who make use of the monsoon currents which reverse every six months, corresponding to the Equatorial counter-current. These could well be Austronesians, if *rati* is a description of outrigger canoes. There is no direct archaeological evidence for such early period contacts, but other indications are highly suggestive. These draw from oral traditions, textual references, maritime technology, plant and animal transfers, disease and other aspects of material culture (e.g. Hornell 1928, 1936, 1941; Grottanelli 1947; Walsh in press). Evidence for this exchange is the transfer of elephantiasis to Africa and the export of African malaria, and the import of SE Asian fighting cocks[6] and bananas, *huti*, to the East African coast (Blench 2009a). So far there is no direct archaeological evidence, but then there was no evidence for Graeco-Roman trade on the coast until the 1990s, despite the unambiguous textual evidence (Juma 1996).

12 West Africa

Postulating direct Austronesian contact with West Africa may seem much more unlikely. This hypothesis has a long and somewhat unhappy history beginning with the arguments of Jones (1971) that the African xylophone is an Indonesian import, a claim discounted in Blench (1982). Nonetheless, there *is* a striking problem of explaining the early presence of certain SE Asian food crops on the West African coast (Blench 2009a). These are: (i) the plantain or triploid banana (AAB); (ii) the water-yam (*Dioscorea esculenta*); and (iii) the cocoyam (*Colocasia esculenta*).

Phytolith evidence places the plantain in Cameroun ca. 500 BC (Mbida et al. 2000). For the other two, the evidence is that their greatest genetic diversity is in West Africa and they are hardly used in East Africa. These species are not native to Africa, cannot have spread across the continent in historical times and are not Portuguese introductions. A piece of contributory evidence is the Nok terracotta statuettes showing elephantiasis, dated to as early as 500 BC, from Central Nigeria (Fagg 1977). Elephantiasis is a Pacific disease that has to be introduced via human migration (Laurence 1968). It remains quite difficult to imagine that stray Austronesian navigators could have rounded the Cape and touched the coast of West Africa more than 2500 years ago, but then many Austronesian voyages would have seemed like impossibilities before they were demonstrated.

13 South Africa

One of the more recent and better documented movements of an Austronesian language to Africa is the evolution and disappearance of Cape Malay. In 1652, employees of the Dutch East India Company moved away from the Cape settlement to clear farms. Since the

[6] Recent research by ILRI has radically revised our understanding of the genetics of the chicken (Han Jianlin pers.comm.). Domestic fowl are now known to have three centres of domestication, India, China and island SE Asia. Many of the chickens of Eastern and Southern Africa are derived from island SE Asia and were *not* introduced via any identified intermediary location.

Dutch government would not permit the enslavement of indigenous people but allowed the importation of slaves or indentured servants from the Dutch East Indies and elsewhere, this became a favoured source. The first Malay slaves arrived in 1657, and Malays came in a steady stream until the nineteenth century, bringing their distinctive culture and cuisine. Despite the relatively simple narrative given on official websites and general histories, the ethnic composition of the imported slaves was quite complex. According to Armstrong and Worden (1979:120–121), Madagascar was the main regional source of Cape slaves during the period 1652–1834, whereas India and Indonesia (chiefly Macassar and Batavia) contributed much smaller numbers. Nonetheless, the ensemble of slaves became known as Cape Malays and formed a subset of the mixed-race category known as 'coloureds'. As the numbers of *Vrijezwarten* (manumitted slaves) increased in the 18th century a synthetic culture evolved. Although a form of Malay, *Melayu*, seems to have been their common language for a period it began to give way to Afrikaans in the nineteenth century and has now been wholly replaced. Nonetheless, it has left significant traces in Afrikaans, among others the Malay word for 'banana', *piesang*. By a strange irony, 'Cape Malay' culture, notably its cuisine and music, are strongly promoted in today's tourist literature, and have persisted and developed, as the language and original ethnic identity of these Austronesian migrants has been submerged.

14 Austronesians in the New World

14.1 California

If Austronesian voyagers could reach Hawai'i presumably they could also reach the west coast of North America. Suggestions of a general relationship between island SE Asia and North America have an old history but recently a more detailed proposal has been advanced for contact between Polynesians and the Chumash Indians in Southern California between 400 and 800 AD. (Jones and Klar 2005; Klar and Jones 2005). This is based on the unique design of their boats, the *tomol* or sewn plank canoe, and equally striking, Polynesian-type compound fish-hooks. These suggestions remain controversial (see Anderson 2006 and response in Jones and Klar 2006, also Arnold 2007). The Chumash language is relatively well-documented but is essentially dead; nonetheless, this is a question that seems as if it ought to be resoluble through archaeological means.

14.2 Austronesians in South America

Polynesian contact with South America has long been the subject of speculation (Rivet 1926; Buck 1938; Heyerdahl 1941, 1950, 1952, 1963, 1964; Jett 1968; Key 1998; Langdon 2001) but accounts of it were more theatre than history (blond, bearded Norwegian against the elements). The model was confused, imagining Amerindians voyaging in the Pacific, despite their known lack of ocean-going craft. A much more credible model would suppose that Polynesians reached the coast of South America, given their proven maritime skills (Lanning 1969). However, all the hooey surrounding such a transoceanic colonisation convinced many prehistorians to set their face against such contact.

Scholarly scepticism probably cracked with clear evidence that the sweet potato had reached Eastern Polynesia in pre-Hispanic times (Green 1998, 2005). The Quechua name, *kumar*, closely resembles the widespread Polynesian term, *kumala*. There is, however, a

chronological problem with this; Fijian *kūmala* is apparently cognate but must presumably be a borrowing. It seems much more credible that Polynesian contact was quite late, perhaps contemporaneous with the settlement of Easter island, itself now redated to 1200 AD (Hunt and Lipo 2006). Heyerdahl (1964) also pointed to a number of South American plants and adventives which appear to be recorded in the earliest accounts of flora, for example, the Mexican poppy, *Argemone mexicana*. More recent research has added the possibility that the bottle gourd (*Lagenaria siceraria*) reached Eastern Polynesia from the New World, while occurrences in Western Polynesia originate in SE Asia (Green 2000). So some of Heyerdahl's evidence may not be completely misconceived, despite his wayward interpretations. Anderson et al. (2007) have considered the evidence for Ecuadorian sailing rafts and raise the possibility that their capacity to reach westward has been underestimated. They point to some very striking coincidences in imagery between Rapa Nui and Punᶏ island birdmen figures. Bellwood and Hiscock (2005) also use the cut stonework and birdman imagery as evidence for possible South American contact with Easter Island.

Other types of biological evidence have now been presented, notably in human genetics, where the presence of the characteristic Polynesian motif, the 9 base pair deletion, on the west coast of South America points to Polynesian presence (Sykes et al. 1995). Another curious piece of evidence is the 'blue-egg' chicken, a variety of fowl encountered by the first explorers in this region. Genetic studies of the indigenous chickens argue that this is likely to be a descendant of the Polynesian chicken (Storey et al. 2007) although Gongora et al. (2008) have vigorously opposed this interpretation.

The early appearance of the coconut in the New World has been the subject of considerable controversy. The coconut was previously considered a New World domesticate that spread westwards across the Pacific, but very early dates for coconut in the Sepik suggest a Malesian domesticate. Nonetheless, Zizumbo-Villareal and Quero (1998), in a re-examination of the earliest sources, argue that it was definitely present on the west coast of Central America in the pre-Hispanic era. Baudouin and Lebrun (2009) examined molecular markers for Central American coconuts and compared them with insular SE Asia. The closest similarities are with those of the Philippines and both are quite distinct from the South American coconut cultivars, suggesting two quite distinct introductions. Furthermore such an origin rules out distribution on ocean currents as far as this can be gauged. Whether the date they attach to this introduction (2250 BP) can be justified is more doubtful, but this presents additional evidence for early trans-Pacific contact and perhaps should be matched with the enigmatic distribution of the blow-gun in south-central America (§15). Another possibility, less well-investigated, is the backstrap loom (Broudy 1979). Technologically, it is very similar to those in SE Asia, and is distributed from Peru through to Central America and the American Southwest. No archaeological finds of textiles apparently using this technology are older than ca. 500 AD, so it has potential for further investigation.

A further interesting thread is the possibility of relatively late contact between the Mapuche Indians of South-Central Chile and Eastern Polynesians (Ramirez 1990/91). Examples of apparent loanwords and other cultural artefacts such as clubs similar to the Maori *patu* make this a possibility, and it certainly has not been rejected out of hand by specialists. However, it would represent a distinct and chronologically different layer from the other contacts discussed in this paper.

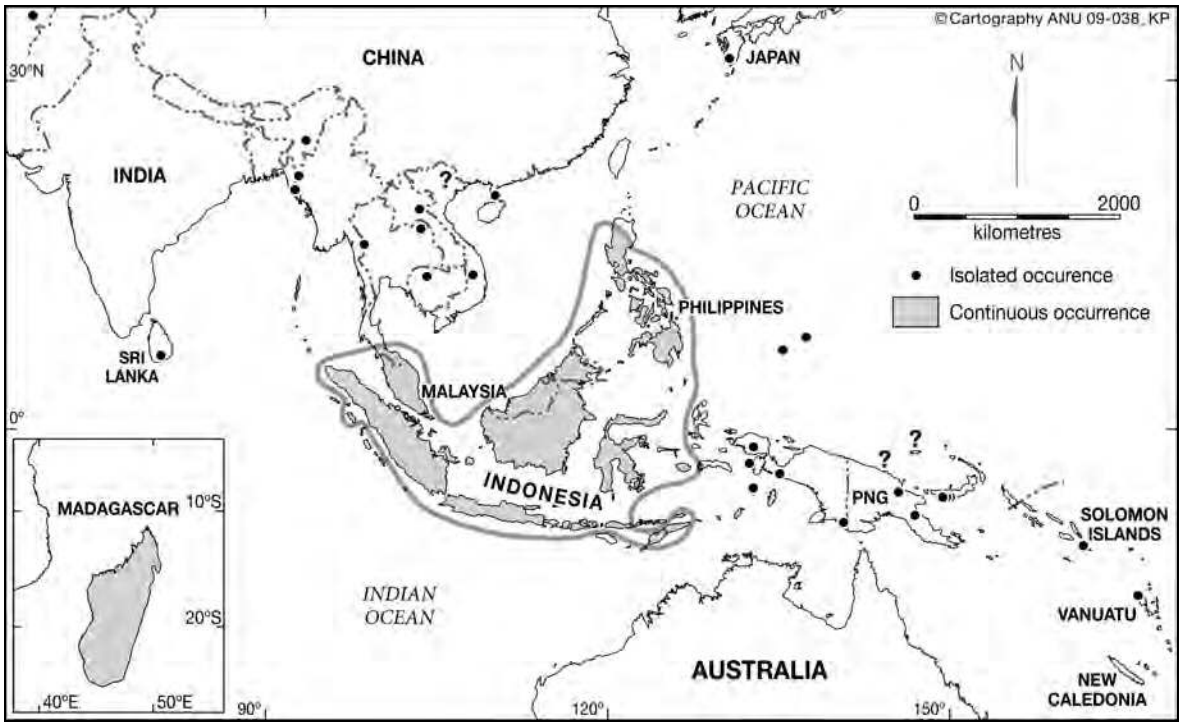
14.3 Surinam

One of the more surprising later adventures of the Austronesians is the presence of a Javanese community in Surinam, the former Dutch colony. The first Javanese came to Surinam in 1890 under slightly confused circumstances (Dew 1981). The settlers multiplied significantly and now represent some 20% of the population. Linguistically, the Javanese spoken in Surinam resembles most closely the central dialects, such as that in Kedu residency. Although the older generation has struggled to maintain cultural traditions, and the usual ethnic revival processes are at work, it seems unlikely that the language is viable in the long term.

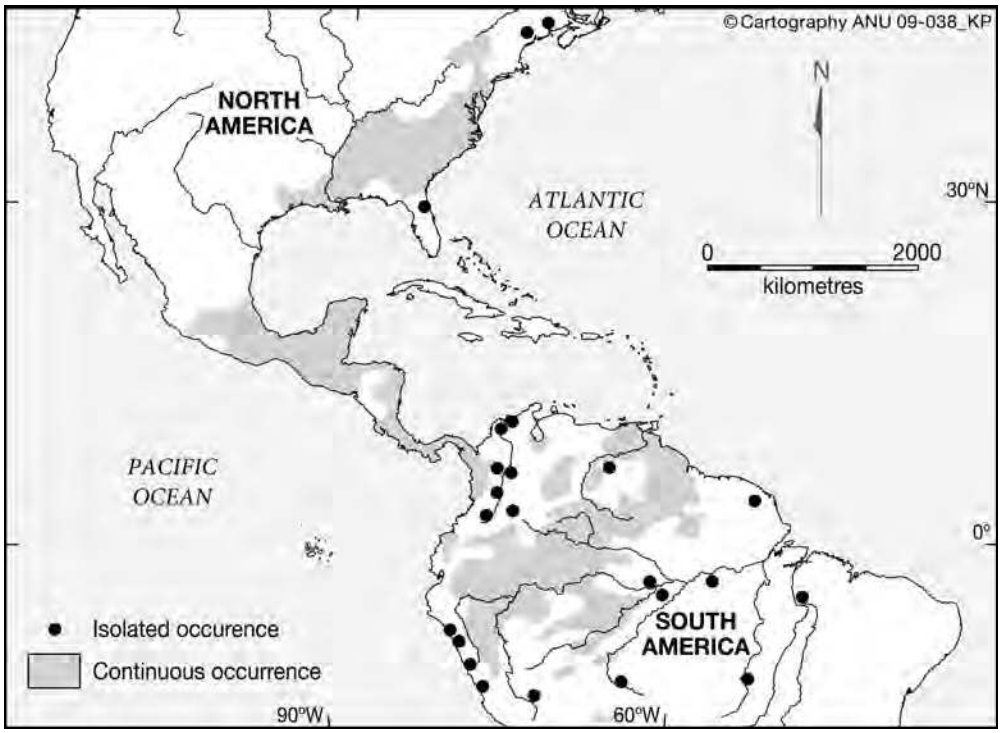
15 The strange case of the blow-gun

There is an old tradition of discussions concerning Old and New World similarities, for example the game parchesi, which is very similar in Asia and North America. The key issue is what level of technological complexity is likely to be subject to convergent invention, as opposed to being evidence for diffusion. Archaeologists typically support a strong 'local invention' emphasis unless contrary evidence is overwhelming. One of the most striking examples of a technology spread by Austronesian contact is the blow-gun. Blow-guns used for hunting or warfare may seem like a technology that could be invented many times, but their worldwide distribution is very patchy. Entirely absent from Africa and Western Eurasia, they are typical of SE Asia and a distinctive zone between North and South America (Jett 1970, 1991). Virtually all occurrences within the Old and New World distributions are contiguous, suggesting the technology was only adopted once and diffused, rather than developing through convergent evolution. In the Old World, the blow-gun maps extremely well against the proposed Austronesian migration sites listed in this paper, primarily in the Austronesian-speaking areas of insular SE Asia and the Pacific (Map 3). The absence of the blow-gun in Taiwan and South China argues that if the case for Daic as an Austronesian branching is correct then it would have occurred prior to the development of the blow-gun in the Northern Philippines. Occurrences in eastern Indo-China probably reflect the Chamic incursions, while those on the Irrawaddy and west Thailand may reflect the exploratory voyages apparently connected with the transfer of ricefield commensals (§5). Occurrences in South India and Sri Lanka might be additional evidence for the contact proposed by Waruno Mahdi (§8).

However, the more controversial issue is whether the distribution of the blowgun in the New World might reflect contact with SE Asia. Essentially, the blowgun occurs in two areas where occurrences are broadly contiguous, Central America and the Amazon and the SE United States (Map 4). The absence of the blowgun over very large areas of the New World and its complete absence in the Centre and Northwest of the United States argues that it cannot be a technology connected with the Bering Strait migrations. Jett (1991) argues convincingly on technological grounds that the distribution in the SE United States is likely to be a sea-borne diffusion from further south (and this is certainly not impossible given recent understanding of early maritime voyaging in the Caribbean). However, the real problem is that if the blow-gun *were* to be due to Austronesian contact it would have to be fairly old, simply because the distribution is so broad. Realistically, it would have to be part of the same expansion that brought early Austronesian landings on the Marianas, i.e. ca. 3500 BP (Butler 1995). The blow-gun is not known on Taiwan, so this would be related to its apparent invention in the Northern Philippines.



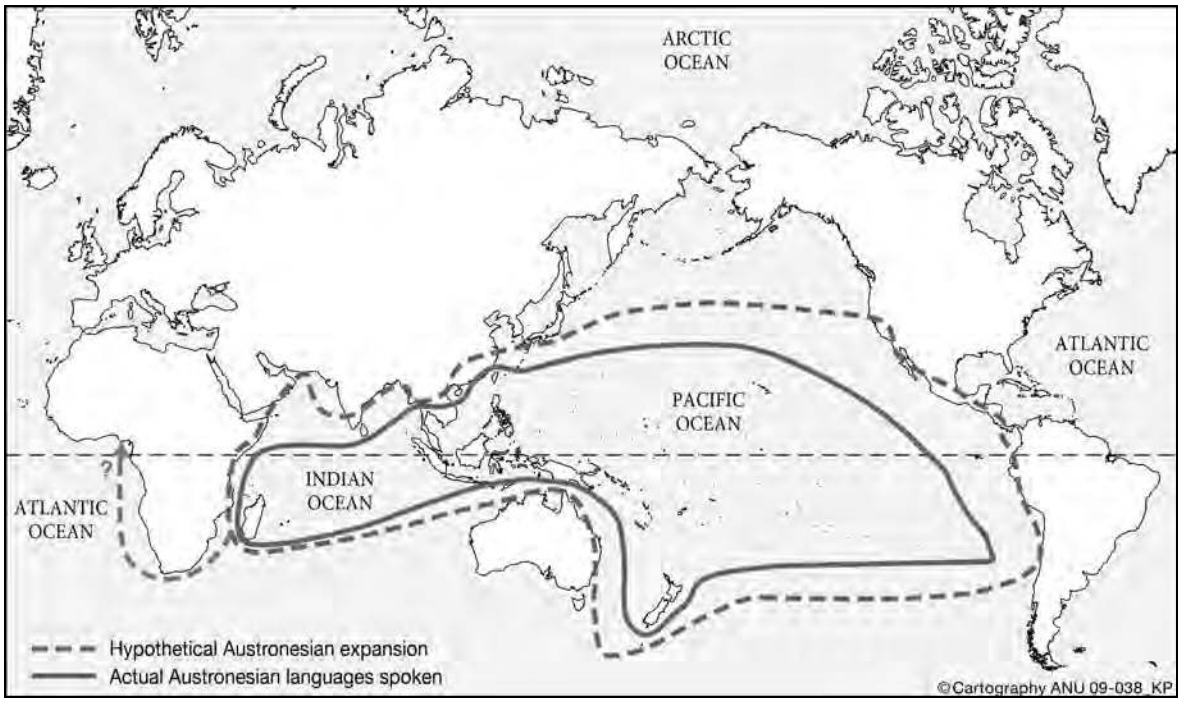
Map 3: Blowgun distribution in the Old World (after Jett 1991)



Map 4: Blowgun distribution in the New World (after Jett 1970)

16 Conclusions

The usual concept of the Austronesian expansion is dominated by the presence of Austronesian languages. The much-reproduced maps show the Pacific and a typical loop westwards to Madagascar. Proposals for the presence of Austronesians elsewhere have a long history and range from the frankly marginal, via the debatable to the likely. The evidence presented in this paper strongly suggests that it is not always helpful to frame our thinking using only synchronic evidence. In some ways this makes the Austronesian adventure even more remarkable, and its exclusion from the mainstream narrative of world history even more inexcusable. It should suggest to linguists that that there is a compelling case for examining more closely the languages in areas where Austronesian was formerly spoken for evidence of substrates or loanwords. Map 5 contrasts the conventional distribution of Austronesian languages with the possibilities discussed in this paper, providing an expanded frame of reference for Austronesianists.



Map 5: Extent of contemporary Austronesian and possible further migrations

References

Adelaar, K.A. 1991. Some notes on the origin of Sri Lankan Malay. In H. Steinhauer, ed. *Papers in Austronesian linguistics No.1*, 23–37. Canberra: Pacific Linguistics.

—— 1996. Malagasy culture-history: some linguistic evidence. In Reade, ed. 1996, 487–500.

—— 2006. The Indonesian migrations to Madagascar: making sense of the multi-disciplinary evidence. In Truman Simanjuntak, Ingrid H.E. Pojoh and Mohammad Hisyam, eds *Austronesian diaspora and the ethnogenesis of people in the Indonesian*

Archipelago. *Proceedings of the International Symposium*, 205–232. Jakarta: LIPI Press (Indonesian Institute of Sciences).

—— in press. Towards an integrated theory about the Indonesian migrations to Madagascar. In P.N. Peregrine, I. Peiros and M. Feldman, eds *Ancient human migrations: a multidisciplinary approach*. Salt Lake City: University of Utah Press.

Anderson, Atholl J. 2000. Slow boats from China: issues in the prehistory of Indo-Pacific seafaring. In S. O'Connor and P. Veth, eds *East of Wallace's Line: studies of past and present maritime cultures of the Indo-Pacific region*, 13–50. Rotterdam: Balkema.

—— 2002. Faunal collapse, landscape change and settlement history in Remote Oceania. *World Archaeology* 33.375–390.

—— 2005. Crossing the Luzon Strait: archaeological chronology in the Batanes Islands, Philippines and the regional sequence of the Neolithic dispersals. *Journal of Austronesian Studies* 1(2).25–44.

—— 2006. Polynesian seafaring and American horizons: a response to Jones and Klar. *American Antiquity* 71.759–763.

Anderson, Atholl J., H. Martinsson-Wallin and Karen Stothert. 2007. Ecuadorian sailing rafts and Oceanic landfalls. In A. Anderson, K. Green and F. Leach, eds *Vastly ingenious: the archaeology of Pacific material culture. In honour of Janet M. Davidson,* 117–134. Dunedin: Otago University Press.

Armstrong, J.C. and N.A. Worden. 1979. The slaves, 1652–1834. In R. Elphick and H. Giliomee, eds *The shaping of South African society, 1652–1840*. Cape Town: Maskew Miller Longman.

Arnold, J.E. 2007. Credit where credit is due: the history of the Chumash oceangoing plank canoe. *American Antiquity* 72.196–209.

Baudouin, Luc and Patricia Lebrun. 2009. Coconut (*Cocos nucifera* L.) DNA studies support the hypothesis of an ancient Austronesian migration from Southeast Asia to America. *Genetic resources of crop evolution* 56.257–262. [Available online: http://www.springerlink.com/content/x7775w7330764347/fulltext.pdf]

Beaujard, Philippe. 2003. Les arrivées austronésiennes à Madagascar: vagues ou continuum? (Partie 1, 2). *Études Océan Indien* 35–36, 59–147.

Bellwood, P. 1978. *Man's conquest of the Pacific*. London: Collins.

—— 2007. Southeast China and the prehistory of the Austronesians. In T. Jiao, ed. *Lost maritime cultures: China and the Pacific*, 36–53. Hawaii: Bishop Museum Press.

Bellwood, P. and P. Hiscock. 2005. Australia and the Austronesians. In C. Scarre, ed. *The human past*, 264–305. London: Thames and Hudson.

Bellwood, P.J., J. Fox and D. Tryon, eds. 1995. *The Austronesians: historical and comparative perspectives*. Canberra: Research School of Pacific and Asian Studies, The Australian National University.

Benedict, P.K. 1942. Thai, Kadai, and Indonesian: a new alignment in Southeastern Asia. *American Anthropologist* 44.576–601.

—— 1975. *Austro-Thai language and culture, with a glossary of roots*. New Haven: Human Relations Area Files Press.

—— 1990. *Japanese/Austro-Tai* (Linguistica Extranea Studia 20). Ann Arbor: Karoma Publishers.

Blench, R.M. 1982. Evidence of the Indonesian origins of certain elements of African culture: with special reference to the arguments of A.M. Jones. *African music* 6(2).81–93.

—— 1996. The ethnographic evidence for long-distance contacts between Oceania and East Africa. In Reade, ed. 1996:461–470.

—— 2007. New palaezoogeographical evidence for the settlement of Madagascar. *Azania XLII*.69–82.

—— 2008a. Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology? In Sanchez-Mazas et al., eds 2008:105–132.

—— 2008b. A history of fruits on the SE Asian mainland. In Toshiki Osada and Akinori Uesugi, eds *Occasional paper 4: linguistics, archaeology and the human past*, 159–178. Kyoto: Indus Project, Research Institute for Humanity and Nature.

—— 2008c. The Prehistory of the Daic (Tai-Kadai) speaking peoples. Paper presented at 12th International conference of European association of Southeast Asian archaeologists. Leiden, The Netherlands.

—— 2009a. *Bananas and plantains in Africa; re-interpreting the linguistic evidence*. Special issue of *Ethnobotany Research and Applications* edited by Edmond De Langhe, Tim Denham and Luc Vrydaghs. 7:363–380.

—— 2009b. *The Austronesians in Madagascar and their interaction with the Bantu of East African coast: surveying the linguistic evidence for domestic and translocated animals*. Philippines Journal of Linguistics, 18(2):18–43. S. Brainard ed. Manila: SIL.

—— in press. New evidence for the Austronesian impact on the East African coast. In Atholl Anderson, ed. *Global origins and the development of seafaring*. Cambridge: Macdonald Institute.

Blust, Robert A. 1996. Austronesian culture history: the window of language. In W.H. Goodenough, ed. *Prehistoric settlement of the Pacific*. *Transactions of the American Philosophical Society* 86(5).28–35. Philadelphia: American Philosophical Society.

—— 2005. The linguistic history of the Philippines: some speculations. In Hsiu-chuan Liao and C.R.G. Rubino, eds *Current issues in Philippines Linguistics and Anthropology. Parangal kay Lawrence A. Reid,* 31–68. Manila: LSP/SIL Philippines.

Broudy, Eric. 1979. *The book of looms*. London: Studio Vista.

Buck, P. [Te Rangi Hiroa]. 1938. *Vikings of the Pacific sunrise*. New York: J.B. Lippincott.

Butler, B. 1995. *Archaeological investigations in the Achugao and Matansa areas of Saipan, Mariana Islands*. Micronesian Archaeological Survey Report Series No.30. Saipan: Division of Historic Preservation.

Denham, Tim, Mark Donohue and Sara Booth. 2009.Horticultural experimentation in northern Australia revisited. *Antiquity* 83(321).634–648.

Dew, Edward. 1981. *The difficult flowering of Surinam: ethnicity and politics in a plural society*. The Hague: Martinus Nijhoff,

Donkin, R.A. 2003. *Between east and west: the Moluccas and the traffic in spices up to the arrival of Europeans*. Darby, PA: Diane publishing Co.

Donohue, Mark and Tim Denham. To appear. Farming and language in island southeast Asia: reframing Austronesian history. *Current Anthropology*.

Driem, G. van. 2005. The contribution of linguistic palaeontology and Austroasiatic. In Sagart et al., eds 2005:77–80.

Edmondson, J.A. and D.B. Solnit, eds. 1988. *Comparative Kadai: linguistic studies beyond Tai*. Dallas: Summer Institute of Linguistics and University of Texas, Arlington.

Fagg, Bernard. 1977. *Nok terracottas*. Lagos: Ethnographica.

Ferrand, Gabriel. 1934. Sayābidja. In M.Th. Houtsma et al., eds *The encyclopaedia of Islām*, vol. 4. 200–202. Leyden/Luzac, London: Brill.

Flood, Josephine. 2006. *The original Australians: story of the Aboriginal people*. Sydney: Allen and Unwin.

Franklin, Donald C. 2003. Morphology and taxonomy of the Top End bamboo *Bambusa arnhemica* F. Muell., a little-known bamboo from northern Australia. *Bamboo science and culture: the journal of the American Bamboo Society* 17(1). 44–54.

Fuller, D., Lin Qin and Emma Harvey. 2008. Evidence for the late onset of agriculture in the Lower Yangtze region and challenges for an archaeobotany of rice. In Sanchez-Mazas et al., eds 2008:40–83.

Ganter, Regina, Julia Martinez and Gary Lee. 2006. *Mixed relations: Asian-Aboriginal contact in north Australia*. Perth: University of Western Australia Press.

Goeje, M.J. de. 1894. De Sajâbidja. *Feestbundel van taal-, letter-, geschied-, en aardrijkskundige bijdragen ter gelegenheid van zijn tachtigsten geboortedag aan Dr. P.J. Veth oud-hoogleraar door enige vrienden en oud-leerlingen aangeboden*. 10–12. Leiden: Brill.

Gongora, Jaime, Nicolas J. Rawlence, Victor A. Mobegi, Han Jianlin, Jose A. Alcalde, Jose T. Matus, Olivier Hanotte, Chris Moran, Jeremy J. Austin, Sean Ulm, Atholl J. Anderson, Greger Larson, and Alan Cooper. 2008. Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *PNAS* 105(30).10308–10313.

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.

Green, R.C. 1998. Rapanui origins prior to European contact – the view from Eastern Polynesia. In P. Vargas Casanova, ed. *Eastern Island and east Polynesian prehistory*, 87–110. Santiago: University of Chile.

—— 2000. A range of disciplines support a dual origin for the bottle gourd in the Pacific. *Journal of the Polynesian Society* 109.191–197.

—— 2005. Sweet potato transfers in Polynesian prehistory. In C. Ballard, P. Brown, R.M. Bourke and T. Harwood, eds *The sweet potato in Oceania: a reappraisal*, 43–62. Sydney: The University of Sydney, University of Pittsburgh.

Grottanelli, V.L. 1947. Asiatic influences on Somali culture. *Ethnos* 4.153–181.

Groves, Colin. 1995. Domesticated and commensal mammals of Austronesia and their histories. In Bellwood et al., eds, 1995:152–163.

Heyerdahl, Thor. 1941. Did Polynesian culture originate in America? *International Science* 1.15–26.

—— 1950. *Kon-Tiki: across the Pacific by raft*. New York: Rand McNally.

—— 1952. *American Indians in the Pacific: the theory behind the Kon-Tiki expedition*. London: George Allen and Unwin.

—— 1963. Feasible ocean routes to and from the Americas in pre-Columbian times. *American Antiquity* 28.482–488.

—— 1964. Plant evidence for contacts with America before Columbus. *Antiquity* 38.120–133.

—— 1986. *The Maldive mystery*. London: George Allen and Unwin.

Hornell, J. 1920. The origins and ethnological significance of Indian boat designs. *Memoirs of the Asiatic Society of Bengal* 7:139–256.

—— 1928. Indonesian culture in east Africa. *Man* 28.1.

—— 1936. Indonesian influence on east African culture. *Journal of the Royal Anthropological Institute* LXIV.305–332.

—— 1941. The sea-going *mtepe* and *dáu* of the Lamu archipelago. *Mariner's Mirror* 27.54–68.

Hudson, Mark J. 1999. *Ruins of identity: ethnogenesis in the Japanese islands*. Honolulu: University of Hawai'i Press.

Hudson, Mark J., Yu Taneishi and Minoru Yoneda. submitted. *New Okinawan Site Marks Northwest Boundary of Prehistoric Austronesian Settlement*.

Hunt, T.L. and C.P. Lipo. 2006. Late colonization of Easter Island. *Science* 311. 1603–1606.

Hunter-Anderson, R.L., G.B. Thompson and D.R. Moore. 1995. Rice as a prehistoric valuable in the Mariana Islands, Micronesia. *Asian perspectives* 34(1).69–89.

Hussainmiya, B.A. 1987. *Lost cousins: the Malays of Sri Lanka*. Occasional Paper, vol. 2. Kebangsaan: Universiti Kebangsan.

Jett, Stephen C. 1968. Malaysia and tropical America: some racial, cultural and ethnobotanical comparisons. *XXXVII Congreso Internacional de Americanistas, Actas y Memorias* 4. 133–177.

—— 1970. The development and distribution of the blowgun. *Annals of the Association of American Geographers* 60(4).662–688.

—— 1991. Further information on the geography of the blowgun and its implications for early transoceanic contacts. *Annals of the Association of American Geographers* 81(1).89–102.

Jones, A.M. 1971 [2nd edition], Africa and Indonesia. *The evidence of the xylophone and other musical and cultural factors.* Leiden: Brill.

Jones, Terry L. and Kathryn A. Klar. 2005. Diffusionism reconsidered: linguistic and archaeological evidence for prehistoric Polynesian contact with southern California. *American Antiquity* 70.457–484.

—— 2006. On open minds and missed marks: a response to Atholl Anderson. *American Antiquity* 71.765–770.

Juma, A. M. 1996. The Swahili and the Mediterranean worlds: pottery of the late Roman period from Zanzibar. *Antiquity* 70.148–54.

Key, Mary Ritchie. 1998. Linguistic similarities between Austronesian and South American Indian languages. *Pre-Columbiana* 1(1/2).59–71.

Klar, Kathryn A. and Terry L. Jones. 2005. Linguistic evidence for a prehistoric Polynesia-southern California contact event. *Anthropological Linguistics* 47.369–400.

Koler-Matznick, Janice. 2002. The origin of the dog revisited. *Anthrozoös* 15(2).98–118.

Kumar, Ann. 1998. An Indonesian Element in the Yayoi? The evidence of biological anthropology. *Anthropological Science* 106(3).265–276.

—— 2007. *Rethinking the prehistory of Japan: Java and the foundation of Japanese society*. London: Kegan Paul.

Kumar, Ann and Phil Rose. 2000. Lexical evidence for early contact between Indonesian languages and Japanese. *Oceanic Linguistics* 39(2).219–255.

Langdon, Robert. 2001. The bamboo raft as a key to the introduction of the sweet potato in prehistoric Polynesia. *The Journal of Pacific History* 36(1).51–76.

Lanning, E.P. 1969. South America as source for aspects of Polynesian cultures. In R.C. Green and M. Kelly, eds *Studies on Oceanic Culture History*, 175–185. Honolulu: Bernice P. Bishop Museum.

Laurence, B.R. 1968. Elephantiasis and Polynesian origins. *Nature* 219.53–56.

Levitt, D. 1981. *Plants and people: Aboriginal uses of plants on Groote Eylandt*. Canberra: Australian Institute of Aboriginal Studies.

Li, Paul Jen-Kuei, Cheng-Hwa Tsang,Ying-Kuei Huang, Dah-An Ho and Chiu-Yu Tseng, eds, 1995. *Austronesian studies relating to Taiwan*. Taipei: Institute of History and Philology, Academia Sinica

Macknight, C.C. 1976. *The voyage to Marege. Macassan trepangers in northern Australia*. Melbourne: Melbourne University Press.

Mahdi, Waruno. 1999a. The dispersal of Austronesian boat forms in the Indian Ocean. In R.M. Blench and M. Spriggs, eds *Archaeology and Language III,* 144–179. London: Routledge.

—— 1999b. Linguistic and philological data towards a chronology of Austronesian activity in India and Sri Lanka. In R.M. Blench and M. Spriggs, eds *Archaeology and Language IV,* 160–242. London: Routledge.

Manguin Pierre-Yves. 1993. Trading ships of the South China Sea. Shipbuilding techniques and their role in the history of the development of Asian trade networks. *Journal of the Economic and Social History of the Orient* 36(3).253–280.

Mbida, C.M., W. Van Neer, H. Doutrelepont and L. Vrydaghs. 2000. Evidence for banana cultivation and animal husbandry during the first millennium BC in the forest of Southern Cameroon. *Journal of Archaeological Science* 27.151–162.

Meacham, W. 1984–85. On the improbability of Austronesian origins in South China, *Asian Perspectives* 26(1).89–106.

—— 1991. Further considerations of the hypothesized Austronesian Neolithic migration from South China to Taiwan and Luzon, *Bulletin of the Indo-Pacific Prehistory Association* 11.398–407.

—— 1995. Austronesian origins and the peopling of Taiwan. In Li et al., eds, 1995:227–253.

—— 2004. Solheim and Austronesian origins. In V. Paz, ed. *Southeast Asian archaeology, Wilhelm G. Solheim II festschrift*, 53–63. Quezon City: University of the Philippines Press.

Moodley, Yoshan, Bodo Linz, Yoshio Yamaoka, Helen M. Windsor, Sebastien Breurec, Jeng-Yih Wu, Ayas Maady, Steffie Bernhöft, Jean-Michel Thiberge, Suparat Phuanukoonnon, Gangolf Jobb, Peter Siba, David Y. Graham, Barry J. Marshall, Mark Achtman. 2009. The Peopling of the Pacific from a Bacterial Perspective. *Science* 323.527–530.

Morinaga, T. 1968. Origin and geographical distribution of Japanese rice. *Japan Agricultural Research Quarterly* 3(2).1–5.

Munoz, Paul Michel. 2006. *Early kingdoms of the Indonesian archipelago and the Malay Peninsula*. Singapore: Editions Didier Millet.

O'Grady, G.N. and D.T. Tryon. 1990. Early Austronesian loans in Pama-Nyungan? In G.N. O'Grady and D.T. Tryon, eds *Studies in comparative Pama-Nyungan*, 105–116. Canberra: Pacific Linguistics.

Oppenheimer, Stephen. 2004. The 'express train from Taiwan to Polynesia': on the congruence of proxy lines of evidence. *World Archaeology* 36(4).591–600.

Ostapirat, Weera. 2005. Kra-Dai and Austronesian: notes on phonological correspondences and vocabulary distribution. In Sagart et al., eds, 2005:107–131.

Pawley, A. 2003. The Austronesian dispersal: languages, technologies and people. In P. Bellwood and C. Renfrew, eds *Examining the farming/language dispersal hypothesis*, 251–273. Cambridge: McDonald Institute for Archaeological Research.

Pawley A.K. and M. Ross. 1993. Austronesian historical linguistics and culture history. *Annual review of anthropology* 22.425–429.

Rackham, H. 1942. *Pliny: Natural History. Books II–VII,* Loeb Classical Library. Cambridge, MA: Harvard University Press.

Ramirez, J.M. 1990/91. Transpacific contact: the Mapuche connection. *Rapa Nui Journal,* 4(4).53–55.

Ray, Himanshu Prabha. 1989. Early maritime contacts between south and southeast Asia. *Journal of Southeast Asian Studies* 20(1).42–54.

—— 2003. *Archaeology of seafaring in ancient South Asia: the formative period*. Cambridge: Cambridge University Press.

Reade, J., ed. 1996 *The Indian Ocean in antiquity*. London/New York: Kegan Paul/British Museum.

Rivet, Paul. 1926. Les Malayo-polynésiens en Amérique. *Journal de la Société des Américanistes de Paris* XVIII.141.

Rolett, B. 2007. Southeast China and the emergence of Austronesian seafaring. In T. Jiao, ed. *Lost maritime cultures: China and the Pacific*, 54–61. Hawaii: Bishop Museum Press.

Rolett, B., T. Jiao and G. Lin. 2002. Early seafaring in the Taiwan Strait and the search for Austronesian origins. *East Asian Archaeology*, 307–319.

Rowland, M.J. 1981. Radiocarbon dates for a shell fishhook and disc from Mazie Bay, North Keppel Island. *Australian Archaeology* 12.63–69.

—— 1982. Keppel Islands marine specialists: an adaptation to the southern Barrier Reef province. In S. Bowdler, ed. *Coastal archaeology in eastern Australia*, 114–120. Canberra: Research School of Pacific Studies, The Australian National University.

—— 1984. A long way in a bark canoe: Aboriginal occupation of the Percy Islands. *Australian Archaeology* 18.17–31.

—— 1987. The distribution of Aboriginal watercraft on the east coast of Queensland: implications for culture contact. *Australian Aboriginal Studies* 2.38–45.

Sagart, Laurent. 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics* 33(2).271–308.

—— 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2).411–444.

—— 2005a. Sino-Tibetan-Austronesian: an updated and improved argument. In Sagart et al., eds 2005:161–176.

—— 2005b. Tai-Kadai as a subgroup of Austronesian. In Sagart et al., eds 2005:177–181.

Sagart, Laurent, Roger Blench and Alicia Sanchez-Mazas, eds. 2005. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London: Routledge Curzon.

Sanchez-Mazas, Alicia, R.M. Blench, M.D. Ross, I. Peiros and Marie Lin, eds. 2008. *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*. London: Routledge.

Solheim, II W.G. 1984–85. The Nusantao hypothesis: the origin and spread of Austronesian speakers. *Asian perspectives* 26(1).77–88.

Storey, Alice A, José Miguel Ramírez, Daniel Quiroz, David V. Burley, David J. Addison, Richard Walter, Atholl J. Anderson, Terry L. Hunt, J. Stephen Athens, Leon Huynen, and Elizabeth A. Matisoo-Smith. 2007. Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *PNAS*, 104(25).10335–10339.

Summerhayes, Glenn R. and Atholl Anderson. in press. An Austronesian presence in southern Japan: early occupation in the Yaeyama islands. *Asian Perspectives.*

Susumu, Ōno. 1970. *The origin of the Japanese language*. Tokyo: Kokusai Bunka Shinkokai.

Sykes, B., A. Leiboff, J. Low-Beer, S. Tetzner and M. Richards. 1995. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics* 57(6).1463–1475.

Szabo, K. and S. O'Connor. 2004. Migration and complexity in Holocene Island southeast Asia. *World Archaeology* 36.621–628.

Terrell, John Edward. 2004. The 'Sleeping Giant' hypothesis and New Guinea's place in the prehistory of Greater Near Oceania. *World Archaeology* 36(4).601–609.

Terrell, John Edward, Kevin M. Kelly and Paul Rainbird. 2001. Foregone conclusions? In search of 'Papuans' and 'Austronesians'. *Current Anthropology* 42(1).97–124.

Thurgood, Graham. 1994. Tai-Kadai and Austronesian: the nature of the relationship. *Oceanic Linguistics* 33(2).345–368.

Tsang Cheng-hwa. 1995. New archaeological data from both sides of the Taiwan straits and their implications for the controversy about Austronesian origins and expansion. In Li, et al., eds 1995:185–225.

—— 2001. Maritime adaptations in prehistoric southeast China: implications for the problem of Austronesian expansion. *Journal of East Asian archaeology* 3(1/2).15–46.

Vovin, Alexander. 1994. Is Japanese related to Austronesian? *Oceanic Linguistics* 33(2).369–390.

Walsh, Martin. in press. Deep memories or symbolic statements? The Diba, Debuli, and related traditions of the east African coast. *Festschrift for Claude Allibert*. Paris: INALCO.

Zizumbo-Villareal, Daniel and Hermilo J. Quero. 1998. Re-evaluation of early observations on coconut in the New World. *Economic Botany* 52(1).68–77.

4 *The historical value of single words*

ROBERT BLUST

1 Introduction[1]

Historical linguists are accustomed to working with large quantities of data, and for this reason it is easy to fall into the habit of believing that any inference of real scientific value must depend on heavy documentation. One thinks in this connection of Brugmann's (1884:253) admonition that linguistic subgroups can only be established safely by demonstrating a mass of exclusively shared innovations. There is a general recognition that the etymology of individual words can shed light on cultural changes, as where English 'pen (writing implement)' and the semantic equivalent in many other languages of Europe derive from words that originally meant 'feather', but despite this observation it is not commonly believed that single words can have much value in such matters as linguistic subgrouping, the determination of language homelands or providing evidence of language levelling, let alone in shedding light on matters outside the field of linguistics itself, as by dating changes in faunal distributions. The purpose of this paper is to draw attention to the potential value of single etymologies in supporting inferences of a type not usually associated with such lines of evidence. All of the material has been published elsewhere, but it has never been brought together in one place for the purpose of demonstrating the value that single words can have in supporting sometimes quite surprising historical inferences.

2 PAn 'nine' a second time

The aboriginal languages of Taiwan stand apart from most Austronesian languages in a number of respects. One of these is the typology of phoneme inventories. Most Austronesian languages in insular Southeast Asia have only one or two fricatives: /s/, or /s/

[1] It is a pleasure to dedicate this paper to Malcolm Ross, who first approached me in 1975 to express an interest in an academic career dedicated to studying the languages of the Pacific. He has come a long way since then, establishing a record of professional achievement in the field of Austronesian linguistics that few can match, as well as making significant contributions to the still nascent field of Papuan linguistics. Malcolm's many contributions have permanently changed the face of our understanding of the languages of western Melanesia, and his forays into higher-level reconstructions of phonology and syntax via the study of the Formosan aboriginal languages stand out as among the most important and detailed efforts in this area. With his vigour and dedication it is expected and hoped that he will be even more productive in retirement than he has been during his remarkable years at ANU.

and /h/. Formosan languages often have more, and in some cases far more than this. Mayrinax Atayal, for example, is said to have six fricative phonemes /š/, /x/, /h/, /β/, /ž/, /ɣ/ (Li 1995), Saisiyat is said to have a similar number /θ/, /ʃ/, /h/, /β/, /ð/, /ɹ/, the last of these described as a voiced retroflex fricative (Li 1978), and Thao has the seven fricatives /f/, /θ/, /s/, /ʃ/, /h/, /ð/, /ɬ/, which are written in a practical orthography as *f*, *c*, *s*, *sh*, *h*, *z* and *lh* (Blust 2003:18). Even languages that have not developed such a rich inventory of fricative phonemes generally retain the contrast of PAn *s and *S as one of the sibilants that differ in place features. One of the consequences that is sometimes observed in a language with more than one type of sibilant is sibilant harmony (SH), or articulatory interference between sibilants that follow one another in close sequence. As noted in Blust (1995), in Thao this is extreme, affecting a much wider class of segments than is conventionally designated by the term 'sibilant', both in the synchronic and the historical phonology. Table 1 shows the regular reflexes of PAn *C, *b, *d, *z, *j, *S and *R in Thao, and the reflexes of these same protophonemes when a fricative is found elsewhere in the word:

Table 1: Sibilant harmony in Thao

PAn	C	b	d	z	j	S	R
Thao (regular)	c	f	s	s	z	sh	lh
Thao (SH)	sh/sh	lh/lh	lh/lh sh/sh	lh/lh	lh/lh	s/s	

In other words, Thao normally reflects PAn *C as /c/ (voiceless interdental fricative), but if a voiceless palatal fricative /sh/ appears elsewhere in the word, *C becomes /sh/; Thao normally reflects PAn *b as /f/, but if there is a voiceless lateral /lh/ elsewhere in the word *b becomes /lh/, and so on. In the case of PAn *d, which normally becomes /s/, the reflex under sibilant harmony is either /lh/ or /sh/ depending on whether an /lh/ or /sh/ is found elsewhere in the word. Examples include the following, where the first form shows the unconditioned reflex of the protophoneme in question and the second shows the reflex with sibilant harmony: *Caŋis > *canit* 'weep, cry', but *CaqiS > *shaqish* 'sew', *batu > *fatu* 'stone', but *baRuj > *falhuz* ~ *lhalhuz* 'Formosan green pigeon', *dapaN > *sapaz* 'sole of the foot', but *daRa > *lhalha* 'Formosan maple' or *daqiS > *shaqish* 'face', *zalan > *saran* 'path, road', but *zaRum > *lhalhum* 'needle', *Sajek > *shazik* 'smell, odor', but *baRuj > *falhulh* 'dove, pigeon', *Sinaw > *shinaw* 'wash (dishes)', but *Sidi > *sisi* 'wild goat, Formosan serow'. It is noteworthy that /lh/ apparently never assimilates to other fricatives, although its presence conditions assimilatory changes in at least /f/, /s/ and /z/. This may be because the historical change *R > /lh/ preceded *b > /f/, *d/z > /s/, and *j > /z/, creating a situation in which newly created fricatives had to adapt to those already well-established in the language. It is also noteworthy that only one siblilant assimilation appears to be possible in any given word, as seen in *baRuj (expected **falhuz) > *falhulh*, not **lhalhulh.

Although synchronic evidence for sibilant assimilation is lacking in other Formosan languages, irregularities in sound change show that a similar pattern once existed in both Paiwan and Saisiyat (Blust 1995:444–445). This brings us back to the historical value of single words. Dyen (1971:34) claimed that Dempwolff's *siwa 'nine', shows irregular reflexes of the initial consonant in Maanyan and Saaroa:

> For the *s reconstructed by Dempwolff there were perhaps two different sibilants indicated by the difference in treatment in Malagasy and Maanyan … Under a hypothesis that the two languages continue the same proto-language, we could avoid the necessity of reconstructing two different proto-phonemes. Since however there is some evidence in Saaroa (a Formosan language) of a similar distinction we make the provisional reconstructions *s_1 and *s_2.

The wording of this passage suggests that three languages (but only two witnesses) show an irregular development of *s in reflexes of Dempwolff's *siwa 'nine'. In fact, although Maanyan has *suei* for expected **hiwe, Malagasy *sivy* 'nine' is regular. Comparative evidence from the Barito languages, which include Maanyan and Malagasy, suggests Proto Barito *suay, and it is not at all clear that this form is a reflex of *siwa. Among the languages mentioned by Dyen this leaves only Saaroa *u-sia* (expected **u-ia) as showing a clearly irregular reflex. However, as noted in Blust (1995:447), at least seven Formosan languages which distinguish *s from *S have a cognate of Dempwolff's *siwa, and these consistently point to *Siwa, not *siwa. The discrepancy in sound correspondence thus comes down to an opposition of the Austronesian languages of Taiwan (which indicate *Siwa), and those outside Taiwan (which indicate *siwa). Without considering any other information it might prove difficult to determine which of these forms is conservative, and which innovative. For those who accept Dempwolff's reconstructions as a fixed standard it would follow automatically that Formosan languages show an unexplained innovation in this form. But it is obvious that Dempwolff's reconstructions, which took no account of Formosan languages, are not a fixed standard for the reconstruction of Proto Austronesian. Since this leaves the direction of change in this form an open question, it becomes necessary to search for a reason why *s would irregularly become *S or vice-versa.

This brings us back in turn to sibilant assimilation. It is well known that the onsets of successive numerals often interfere with one another in language history, as with English four (expected **whour), five. Bloomfield (1933) cites a number of examples of such assimilatory irregularities in other Indo-European languages, including Latin, Greek, and Old Bulgarian, and labels this general phenomenon 'contamination'. In each of these cases successive numerals that conformed to regular historical changes would have had phonetically similar, but non-identical onsets. Instead, they show identical onsets as a result of either anticipatory or perseverative assimilation. Comparable examples of contamination affecting the onsets of successive numerals in many Tibeto-Burman languages are cited by Matisoff (1995). It is known that the PAn base for 'ten' was *puluq, but that this morpheme actually meant 'group of ten', and acquired the specific meaning 'ten' only when it was preceded by a proclitic form of PAn *esa 'one', hence *sa-puluq. It is also generally assumed that *s and *S were sibilants that differed only in place. If the PAn sequence 9, 10 had been expressed as *siwa, *sa-puluq, in accordance with the reconstructions adopted by Dempwolff, there is no obvious basis for the change PAn *siwa > *Siwa in a hypothetical 'Proto Formosan'. On the other hand, if the PAn sequence 9, 10 was *Siwa, *sa-puluq, the conditions for sibilant assimilation and hence contamination between the onsets of successive numerals would have existed, leading to a lexically specific phonological innovation PAn *Siwa > *siwa. Since all languages outside Taiwan which have unambiguous reflexes of this form point to *siwa, the history of this single word turns out to have subgrouping implications of an order far greater than one would expect on a priori grounds. It is true that other lines of evidence support a non-Formosan, or Malayo-Polynesian subgroup, but the power of this etymology is such that even if no other evidence were known, we would still be forced to seriously consider the Malayo-

Polynesian hypothesis on the basis of the irregularity in this single word, which must have happened shortly before the change *S > *h, since the latter is reflected in all Malayo-Polynesian languages.[2]

3 Chamorro *pakyo* 'typhoon'

The position of Chamorro within the Austronesian language family has long challenged historical linguists. The general view is that it is a Malayo-Polynesian language that has no close relatives. Three specific ideas that have been advanced in recent years are: (i) that Chamorro subgroups with languages of the Philippines, and is most closely related to Ilokano and Tagalog (Topping 1973:3); (ii) that Chamorro reached the Mariana islands through a migration directly from Taiwan (suggested in Starosta and Pagotto 1991:345–346, and then advanced more vigorously in Starosta 1995:694–695); and (iii) that Chamorro reached the Mariana islands from Sulawesi in central Indonesia (Zobel 2002:432). Topping's view, which was not formally argued, was based on a superficial consideration of the voice system. The latter two views rely on grammatical arguments, but neither is directly concerned with other lines of evidence. Starosta (1995) effectively claimed that Chamorro is descended from a putative 'Proto Formosan' for which no convincing evidence of any kind exists, and that it is more closely related to such Formosan aboriginal languages as Kanakanabu, Bunun, Seediq, Atayal, Saisiyat, Paiwan and Amis than it is to Malayo-Polynesian languages such as Tagalog or Ilokano. Zobel, on the other hand, places Chamorro directly under a 'Nuclear Malayo-Polynesian' node which includes languages of western Indonesia, Chamorro, Palauan and Central-Eastern Malayo-Polynesian languages, but not languages of the Philippines, north Sulawesi, or northeast and interior Borneo.

To the extent that the Malayo-Polynesian hypothesis is valid, it follows that Chamorro is a Malayo-Polynesian language. This effectively undercuts any claim that Chamorro reached the Marianas directly from Taiwan, since if that were the case it would become necessary to attribute a number of phonological, lexical and grammatical innovations that it shares with other Austronesian languages outside Taiwan to massive unmotivated convergence. This leaves the Philippines and Sulawesi as possible homelands that have been proposed for the pre-Chamorro linguistic community. Based on four proposed morphosyntactic innovations, Zobel (2002:432) suggests that both Chamorro and Palauan migrated into the Pacific directly from Sulawesi: 'It was probably also from Sulawesi that the speakers of Chamorro and Palauan (or better: Pre-Chamorro and Pre-Palauan) sailed to the northeast to the distant islands of Palau and the Marianas.' There is, however, no phonological or lexical support for such an inference. This is an important observation, since one particular etymology has far greater than ordinary bearing on the question of the pre-Chamorro homeland.

PAn *baRiuS 'typhoon' can be reconstructed on the basis of cognate forms in Formosan and non-Formosan languages such as Saisiyat *baLyosh*, Amis *faliyus*, Puyuma *vaRiw*, Tagalog *bagyó*, Cebuano *bagyú* 'typhoon'. The typhoon zone extends from a region to the east of Chuuk, at about 7 degrees north latitude, westward and slightly northward to the Philippines, Taiwan, and southern Japan. As noted in Blust (2000) there is no evidence that Chamorro belongs to the Philippine subgroup of Austronesian languages.

[2] Although some historical linguists, as Lehmann (1992:224) and Campbell (2004:118–120) consider contamination to be a type of blending, the contamination of numeral onsets appears to be essentially an assimilatory process that is unusual in crossing word boundaries.

At the same time Chamorro *pakyo* ‘typhoon’ seems clearly to be a native word (Guam and Saipan lie squarely in the typhoon zone, and there would be no reason to borrow a foreign word for a weather phenomenon that is central to the economic and social life of the people in this region). Typhoon records for the past century in the Philippines indicate that Mindanao is only rarely struck, and regions further to the south do not experience true typhoons at all. Reflexes of *baRiuS nonetheless occur further south, where they refer to strong winds, or to the wind in general, as in Bario Kelabit *bariw*, Kayan *bahuy* ‘strong wind, storm’, Samal *baliw*, Miri *baruy* ‘wind’.

The Chamorro word for ‘wind’ is *maŋlo?*, a form with no known etymology. If the ancestral Chamorro had come from any region south of the Philippines, and very likely any region south of Mindanao, they would have either lost the word for ‘typhoon’, or would have altered it to mean ‘wind’ or ‘strong wind’, and the retained word would then have been semantically altered back to the meaning ‘typhoon’ upon the settlement of the Marianas. While this is not impossible, it is clearly simpler to assume that the word never lost its original sense. However, if this was the case, the evidence suggests that the Marianas were settled from an area in insular Southeast Asia that lies within the western Pacific typhoon zone. This makes Mindanao improbable, and rules out areas further to the south. Since Chamorro is a Malayo-Polynesian language it also rules out Taiwan, leaving the central and northern Philippines as the most plausible source areas. While this is an interesting conclusion in itself, it becomes more complex and intriguing when we recall that there is no evidence for including Chamorro within the Philippine subgroup of Austronesian languages. If the ancestral Chamorros reached the Marianas directly from the central or northern Philippines but Chamorro is not a Philippine language, it follows that the linguistic situation in the Philippines today does not reflect the diversity of Austronesian languages that must have existed at the time the Marianas were settled (now generally agreed to have been by at least 3500 BP). As it happens, this conclusion also follows from the far lower than expected linguistic diversity in the Philippines, an area that apparently has been settled by Austronesian speakers longer than any other region in Southeast Asia apart from Taiwan (Blust 2005). The single Chamorro word *pakyo* ‘typhoon’, then, provides valuable information both about the likely source area in insular Southeast Asia from which the Marianas were settled, and about a likely language levelling event of major proportions which was caused by one early Austronesian language (Proto Philippines) expanding at the expense of others during the Neolithic history of the Philippine islands.

4 PAn *buqaya ‘crocodile’

The saltwater crocodile (*Crocodilus porosus*) is the major large predator in riverine and estuarial environments from northern Australia through the northern Philippines, and from the Asian mainland to the southeast Solomons. East of the Solomons and in most of Micronesia breeding populations of this animal do not exist in a natural state, although individual animals are occasionally found alive or dead.

Reflexes of PMP *buqaya ‘crocodile’ are remarkably stable in languages reaching from the northern Philippines (Itbayaten *vwaaya*, Ilokano *buáya*) to the southeast Solomons (Arosi *huasa*, ‘Āre’āre *huara*). Despite the relatively short distance between the Batanes islands and southern Taiwan, the saltwater crocodile is not found in the latter island. This is surprising, given the presence of a crocodilian species (the Chinese alligator) in the lower Yangzi river, considerably further to the north. What makes this zoogeographical

fact even more interesting is that a single Formosan language has an apparent reflex of *buqaya, namely Puyuma *buaya* 'shark' (Cauquelin 1991) which, if accepted, will raise the status of this reconstruction from PMP to PAn.

In order to accept the etymology PAn *buqaya 'crocodile' > Puyuma *buaya* 'shark' it is necessary to see the semantic difference between reconstruction and reflex as resulting from a plausible change. Since the saltwater crocodile is the major animal threat to humans in estuarial and riverine environments, while the shark occupies this position in purely marine environments, the connection of meaning can be seen as having a strong basis in experience. Moreover, as a direct consequence of the fear felt toward crocodiles and sharks, both animals have become objects of respect or even veneration in various Austronesian-speaking societies. Despite this parallelism between crocodiles and sharks in both practical reality and cultural attitude, the terms used to designate these animals are not known to be interchanged anywhere outside Taiwan. Why, then, would this happen among the Puyuma of southern Taiwan?

The simplest explanation for this semantic shift appears to be that crocodiles were still found in at least southern Taiwan at the time that Austronesian-speaking peoples arrived on the island, and presumably for some period of time thereafter. Darlington (1980:229), citing Pope (1935), notes that the Chinese alligator, which is now restricted to the lower Yangzi river, 'probably … ranged widely in eastern China not long ago.' It is not clear whether the contraction of the earlier natural range of the Chinese alligator is a consequence of human predation, or of other factors, but it is notable that its territory apparently has been significantly reduced over a comparatively short time period. The etymology of Puyuma *buaya* 'shark' suggests a similarly rapid territorial contraction. It appears very unlikely that this word was borrowed from a Philippine language and then applied to a familiar animal, the shark. It is also clear that PAn *buqaya did not mean 'shark', since a better candidate is available for that meaning, namely *qiSu (cf. Paiwan *qisu*, Bikol *ího*, Cebuano *íhu*, Malay, Nias, Sundanese *hiu*, Ngadha, Kambera, Tetun, Leti *iu* 'shark'). The only conclusion left is that PAn *buqaya meant 'crocodile', and that crocodiles were therefore still present in Taiwan at the time of Austronesian settlement, and for some time thereafter. When the territorial range of the saltwater crocodile began to contract, leading to the disappearance of breeding populations of these animals in Taiwan, reflexes of PAn *buqaya were lost in most Formosan languages along with the referent. In Puyuma, however, the term survived by semantic transfer to the only other large and dangerous predator to occupy a similar natural and cultural niche in relation to the human population.

What is perhaps most noteworthy about this linguistic comparison is that it appears to provide evidence about a biological event rather than a linguistic event, namely a contraction of the range of the saltwater crocodile within the relatively recent human past, removing Taiwan from its earlier territorial range.

5 PAn *qaRem 'pangolin'

Two species of pangolin are found in insular Southeast Asia: *Manis pentadactyla* in Taiwan and *Manis javanica* in the Greater Sunda islands and Malay peninsula. Although these animals are distinguishable to a zoologist, they are similar enough in external appearance to make it very likely that if they co-occurred in the same region they would be treated as identical in a folk classification. No members of either species are known in the intervening Philippine islands. A similar distributional discontinuity is seen with other

mammals, including the Asian sun bear, the clouded leopard, and the river otter. From a biological standpoint these mammalian distributions appear similar, in that all of them encompass discontinuous areas, but from a linguistic standpoint they are strikingly different.

Agreements among the Formosan languages point clearly to PAn *Cumay 'bear', *lukeNaw 'clouded leopard' and *Sanaq 'river otter'. Since their referents are unknown in the Philippines, none of these terms appears in Philippine languages. A number of languages in the Greater Sunda islands of western Indonesia, reflect *biRuaŋ 'Malayan sun bear', but reconstructed terms for 'clouded leopard' and 'river otter' have been harder to establish. The point to note in these comparisons is that they do not comprise cognate sets shared by Formosan languages and languages of the Greater Sunda islands. This is what might be expected if there had been a historical discontinuity in contact with the sun bear, clouded leopard and river otter during the Austronesian expansion southward from Taiwan. Given the historically attested distributions of these animals it is easy to fall into the error of assuming a similar history for the pangolin, but here we are surprised to find that the names for the *Manis pentadactyla* and *Manis javanica* are cognate, as reflexes of PAn *qaRem 'pangolin: *Manis pentadactyla*' are widespread both in Taiwan and in Borneo: Atayal *qom*, Seediq *ʔaruŋ*, Saisiyat (Taai) *ʔœLəm,* Pazeh *axem*, Bunun *qalum*, Thao *qalhum*, Saaroa *ʔarəmə*, Kavalan *iRem*, Amis *qalem*, Puyuma *qarem*, Paiwan *qam* 'pangolin: *Manis pentadactyla*', Long Anap Kenyah, Mukah Melanau *aam*, Kelabit, Kiput *arem*, Long Jegan Berawan *akem*, Katingan *ahem*, Maanyan *ayem* 'pangolin: *Manis javanica*'. Given this comparison there is no alternative to reconstructing PAn *qaRem, and to concluding that this term was somehow retained during the Austronesian movement through the Philippines, where the animal is absent. Although nineteenth century reports indicate that *Manis javanica* was present in Palawan and the Calamian and Cuyo islands (which sit on the same Sunda Shelf that underlies Borneo and the other Greater Sunda islands of western Indonesia), there is no evidence that any pangolin species was present in the rest of the Philippines when Austronesian speakers arrived. How, then, was a reflex of *qaRem retained in languages descended from PAn/PMP and re-applied to the newly discovered *Manis javanica* when speakers of Austronesian languages arrived in Borneo?

The most straightforward answer to this question would be to propose that *Manis javanica* was more widely distributed in the Philippines than Palawan and the Calamian and Cuyo islands when Austronesian speakers arrived. If that is the case, however, it is strange that zoologists have found no evidence for earlier presence of this animal elsewhere in the Philippines (Darlington 1980:384). Superficially, the problem raised by the distribution of reflexes of PAn *qaRem 'pangolin' thus resembles the problem raised by the distribution of reflexes of PAn *buqaya 'crocodile'. However, while a plausible case apparently can be made for the presence of crocodiles in at least southern Taiwan after the arrival of Austronesian speakers, a similar case apparently cannot be made for the presence of pangolins in the Philippines outside those islands that rest on the Sunda Shelf. What historical information, then, does the history of this single word provide? If the earlier presence of *Manis javanica* in Luzon, Mindoro, the Bisayas and Mindanao can be decisively ruled out due to the absence of fossil evidence, it would appear that the word for the pangolin was retained for some generations after contact with the animal was lost. One way that this might have been possible is through myths or stories that kept the name of the pangolin alive in the minds of speakers for some generations until they encountered a close approximation of the same animal once again. A second possibility is that Austronesian

speakers moved very rapidly through the Philippines to Borneo, but the linguistic diversity that resulted from this rapid migration was subsequently compromised by the expansion of Proto Philippines at the expense of other early Austronesian languages which were absorbed by it. In either case it is surprising that only the PAn word for 'pangolin' was able to survive the movement of language groups across this apparent distributional hiatus of faunal forms, while the words for 'bear', 'clouded leopard' and 'river otter' did not.

6 Unfamiliar marsupials

Linguistic subgroups are normally established by a consideration of comparative evidence that allows for a clear distinction to be drawn between innovations and retentions (the failure to meet this basic consideration has been the shortcoming of lexicostatistical classifications, which may yield valid or invalid results, depending upon how seriously crosslinguistic variation in the retention rate of basic vocabulary distorts true subgrouping connections). Probably nowhere else in the world, however, has a major zoogeographical boundary been able to play such a key role in determining major linguistic subgroups as in the Austronesian language family.

To biologists concerned with the geographical distribution of animals Indonesia is famous as the site of the Wallace Line, a major zoogeographical boundary that was first described by the English naturalist Alfred Russel Wallace in 1869 (Wallace 1962). Wallace (1962:11) noted that in crossing from Bali to Lombok, two islands just to the east of Java, he was immediately struck by the fundamental differences in land animals and non-migratory birds:

> In Bali we have barbets, fruit-thrushes, and woodpeckers; on passing over to Lombock these are seen no more, but we have abundance of cockatoos, honeysuckers, and brush-turkeys, which are equally unknown in Bali, or any island further west. The strait is here fifteen miles wide, so that we may pass in two hours from one great division of the earth to another, differing as essentially in their animal life as Europe does from America.

At first, the possibility that such a biological division could have any bearing on linguistic subgrouping is not obvious. A further consideration of the evidence, however, reveals a remarkable relationship between linguistic observations and the Wallace Line as it relates to mammalian fauna.

Wallace noted that the division which has since been named after him separates placental mammals on the west (elephants, tigers, bears, pangolins, monkeys, etc.) from marsupial mammals on the east (the cuscus, or phalanger, the bandicoot or marsupial rat, the tree kangaroo, etc.). A few placental mammals evidently crossed this zoogeographic divide early (bats, because of their powers of flight, and rats because of their ability to ride on rafts of vegetation and the like over considerable distances). Apart from these minor exceptions, and domesticated animals that were transported across the Wallace Line by humans (most notably dogs and pigs), the complementation of placental and marsupial mammals is nearly perfect.

For a considerable period of time the position of the languages of eastern Indonesia was in dispute: do these languages subgroup with the Oceanic languages to the east, or with the languages of western Indonesia and the Philippines? Blust (1974, 1978, 1983/84) and Dyen (1978) took opposed stances regarding this question, the former arguing for a connection with Oceanic, and the latter for a connection with languages further to the west. At the time these positions were first adopted nothing was known about the history of

terms for marsupial mammals in Austronesian languages. As it happens, however, these terms provide critical evidence for a major division of the Austronesian language family.

If the Austronesian diaspora began in Taiwan speakers of Austronesian languages would have encountered marsupial mammals for the first time when they crossed the Wallace Line. Under these circumstances the appearance of cognate innovations for these previously unknown animals could have only one interpretation: that the languages in which they are found are descended from a common protolanguage that must have existed when the Wallace Line was crossed. Any other interpretation would force one to the absurd conclusion that speakers of Austronesian languages who crossed the Wallace Line as a result of historically independent migrations had somehow innovated related terms for fauna that were entirely novel to them. The appearance of cognate words for 'cuscus' in both the Moluccas (Watubela *kadola*, Buli *do*) and in Melanesia (Manam *ʔodora*, Vitu *hadora*, Penchal *kotay*, Mussau *aroa*, Nggela *kandora*) can only be explained plausibly by a hypothesis that the languages of eastern Indonesia descend from a protolanguage that is immediately ancestral to them and the Oceanic group (now called 'Central-Eastern Malayo-Polynesian). In this case we are doubly fortunate in being able to extract a historical value from single words, since a cognate set meaning 'bandicoot' also spans many of the languages of eastern Indonesia (Leti-Moa *mada/made*, Damar *madar*, Yamdena *mande*, Amblau *mate*, Asilulu *marel*), and western Melanesia (Motu *mada*, Takia *madal*, Wogeo *m^{w}aja*, Lou *m^{w}as*, Mangap-Mbula *moozo* 'bandicoot', Waidina Fijian *ŋ^{w}*aco, Namosi Fijian *ŋ^{w}aca* 'large rat'). Together these forms point to PCEMP *kandoRa 'cuscus', and PCEMP *mansər/mansar 'bandicoot', and even if no other evidence were available, these two words would provide powerful support for the Central-Eastern Malayo-Polynesian hypothesis.

7 Conclusions

The etymologies of many other single words provide important information on culture history. Examples include PMP *liaŋ 'cave', which has come to refer to any type of burial structure in many of the languages of northern Sarawak, implying the earlier practice of cave burial in that area among peoples who have not practiced it within the ethnographic present, reflexes of *tina mate 'orphan' (lit. 'mother dead') in languages of the Southeast Solomon islands that now have a patrilineal social organisation, implying that they were matrilineal at an earlier time, and the cross-sibling substitution drifts, whereby PMP *ñaRa 'brother (woman speaking)' and *betaw 'sister (man speaking)' were replaced repeatedly in the history of Austronesian languages by terms that have the structural form 'male' or 'male + child' and 'female' or 'female + child' respectively, implying a transfer of terminology from wife-giving and wife-taking lineages to cross-siblings, and hence a system of asymmetric exchange as the basis of PMP social organisation (Blust 1990). The historical value of these single-word comparisons does not differ in kind from that of such well-known examples as English *salary* (< Latin *salārium* 'salt-money'), *pen* (< Latin *penna* 'feather') or clock (< Anglo Saxon *clugge* 'bell'). What is different about the comparisons that are highlighted in this paper is that they shed light on major subgrouping problems (PAn *Siwa > PMP *siwa 'nine', and the PCEMP lexical innovations *kandoRa 'cuscus' and *mansər/mansar 'bandicoot'), on language homelands and important episodes of prehistoric linguistic levelling (Chamorro *pakyo* 'typhoon'), on the contraction of faunal distribution areas (*buqaya 'crocodile'), and somewhat less clearly, on the possibilities of preserving lexical items for generations after losing contact with their referents, and then

revivifying them when similar referents are encountered once again (*qaRem ‘pangolin’). Although all of these conclusions are subject to interpretations which may differ from one researcher to the next, the examples on which they are based are a reminder that the vocabulary of a language is a repository of history that can be mined for far more insights than we are generally apt to notice or appreciate.

References

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.

Blust, Robert. 1974. Eastern Austronesian: a note. *Working Papers in Linguistics* 6(4). 101–107. Honolulu: Department of Linguistics, University of Hawaii.

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

—— 1983/84. More on the position of the languages of eastern Indonesia. Oceanic Linguistics 22–23.1–28.

—— 1990. Austronesian sibling terms and culture history. In Andrew Pawley and Malcolm Ross, eds *Austronesian terminologies: continuity and change*, 31–72. Canberra: Pacific Linguistics.

—— 1995. Sibilant assimilation and the Proto-Austronesian word for ‘nine’: a discourse on method. *Oceanic Linguistics* 34.443–453.

—— 2000. Chamorro historical phonology. *Oceanic Linguistics* 39.83–122.

—— 2003. *Thao dictionary*. Language and Linguistics Monograph Series No. A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

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

Brugmann, Karl. 1884. Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen. *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1.226–256.

Campbell, Lyle. 2004 [1998]. 2nd edition. *Historical linguistics: an introduction*. Cambridge, Massachusetts: The MIT Press.

Cauquelin, Josiane. 1991. *Dictionnaire Puyuma-Français*. Paris/Jakarta: École Français d’Extrême-Orient.

Darlington, Philip J., Jr. 1980 [1957]. *Zoogeography: the geographical distribution of animals*. New York: Robert E. Krieger Publishing Co.

Dyen, Isidore. 1971. The Austronesian languages and Proto-Austronesian. In Thomas A. Sebeok, ed. *Current Trends in Linguistics,* vol. 8: *Linguistics in Oceania*, 5–54. The Hague: Mouton.

—— 1978. The position of the languages of eastern Indonesia. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: Proceedings, Fascicle 1*, 235–254. Canberra: Pacific Linguistics.

Lehmann, Winfred P. 1992. 3rd edition. *Historical linguistics: an introduction*. London and New York: Routledge.

Li, Paul Jen-kuei. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 49(2).133–199.

—— 1995. Atayal. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary, an introduction to Austronesian studies*, Part 1, Fascicle 1, 283–292. Berlin: Mouton de Gruyter.

Matisoff, James A. 1995. Sino-Tibetan numerals and the play of prefixes. *Bulletin of the National Museum of Ethnology* (Osaka) 20(1).105–252.

Pope, C.H. 1935. *The reptiles of China*: *turtles, crocodilians, snakes, lizards.* New York: American Museum of Central History.

Starosta, Stanley. 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, eds *Austronesian studies relating to Taiwan*, 683–726. Symposium Series of the Institute of History and Philology, Academia Sinica, No.3. Taipei: Academia Sinica.

Starosta, Stanley and Louise Pagotto. 1991. The grammatical genealogy of Chamorro. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages. Papers from the Fifth International Conference on Austronesian Linguistics, Parts One and Two*, 319–348. Auckland: Linguistic Society of New Zealand.

Topping, Donald M. 1973. *Chamorro reference grammar*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Wallace, Alfred Russel. 1962 [1869]. *The Malay archipelago*. New York: Dover Publications, Inc.

Zobel, Eric. 2002. The position of Chamorro and Palauan in the Austronesian family tree: evidence from verb morphosyntax. In Fay Wouk and Malcolm Ross, eds *The history and typology of Western Austronesian voice systems*, 405–434. Canberra: Pacific Linguistics.

5 *Beyond pronouns: further evidence for South Bougainville*

BETHWYN EVANS

1 Introduction[1]

The historical relations of the Papuan languages which are scattered across the island region to the east of the New Guinea mainland, including the Bismarck Archipelago in the northwest, Bougainville and the northwest and central Solomon Islands in the east, and Rossel Island in the southwest, remain debated in the linguistic literature.[2]

Ross (2001, 2005) proposes five distinct families of Papuan languages in this region of Island Melanesia, including: (i) Central Solomons; (ii) Yele-West New Britain; (iii) East New Britain; (iv) South Bougainville; and (v) North Bougainville, as well as three language isolates. Ross' evidence for this classification is shared pronominal paradigms, which can be taken as indicative of genetic relationships amongst languages, and that warrant further, and more detailed, investigation.[3]

There have also been proposals of historical connections amongst all the Papuan languages of Island Melanesia. Wurm (1975, 1982) suggests that all these Papuan languages are genetically related, forming an East Papuan Phylum. Wurm states that the evidence for East Papuan consists of both shared similarities in lexicon and typological characteristics,

[1] It is with pleasure that I offer this paper, my first foray into Papuan historical linguistics, to Malcolm Ross; and also gratefully acknowledge the many discussions I have had with Malcolm regarding historical linguistics over the years, as well as those more recently on the data and issues presented in this paper. I would also like to thank Malcolm, Andrew Pawley and Darrell Tryon for providing me with unpublished data on the Papuan and Austronesian languages of Bougainville; and Andrew Pawley and Brigitte Pakendorf for comments on an earlier version of the paper.

[2] The term 'Papuan' is used for languages of the New Guinea mainland and islands to the west in eastern Indonesia and to the east in the islands of Papua New Guinea and the Solomon Islands as a cover term for languages that are not part of the large Austronesian family. However, this label makes no claims about the genetic relatedness of these languages.

Until recently discussions of Papuan languages within this region also included the languages of the Reefs and Santa Cruz islands in eastern Solomon Islands. However, these languages have since been shown to be Austronesian (Ross and Næss 2007).

[3] The problems with using pronouns to establish genetic relatedness have been widely discussed. However, as Ross (2005) describes shared form-meaning pairs in all three person categories are unlikely to be due to chance, but rather are indicative of an historical relationship amongst the languages. Since reported cases of 'borrowed' pronouns tend to be in specific sociocultural contexts and with individual pronominal forms rather than entire paradigms, it is likely that shared paradigms of pronouns reflect a common source (Ross 2005).

but he only gives details of shared structural features and the true diagnostic value of this evidence can be questioned. For example, while systems of gender or nominal classification are widespread in Wurm's 'East Papuan' languages, Terrill (2002) shows that phonological and structural similarities across all the languages, which would be indicative of genetic relatedness, do not exist.

Given that Greater Bougainville was settled about 30,000 years ago (Spriggs 1997), and New Britain and New Ireland about 40,000 years ago (Summerhayes 2007), it is likely that any historical relationship amongst the contemporary Papuan languages of different island groups of Island Melanesia is too distant to be determined by traditional methods of historical linguistics. For this reason, Dunn et al. (2007) and Dunn et al. (2008) investigate genetic relatedness by applying a variety of computational methods derived from evolutionary biology to a database of structural features. They conclude that the linguistic evidence does suggest a historical connection amongst the Papuan languages of this region which likely results from 'either ancient Papuan-Papuan contact, or descent from a common ancestor' (Dunn et al. 2008:748).

This paper investigates a single group of Papuan languages, namely South Bougainville. The paper builds on the work of Ross (2001, 2005) by applying the Comparative Method to a sample of lexical data, and argues that the regular sound correspondences, which can be established on the basis of form-meaning pairings in basic vocabulary, are additional support for the genetic relatedness of the four Papuan languages of southern Bougainville.

2 Further evidence for a South Bougainville family

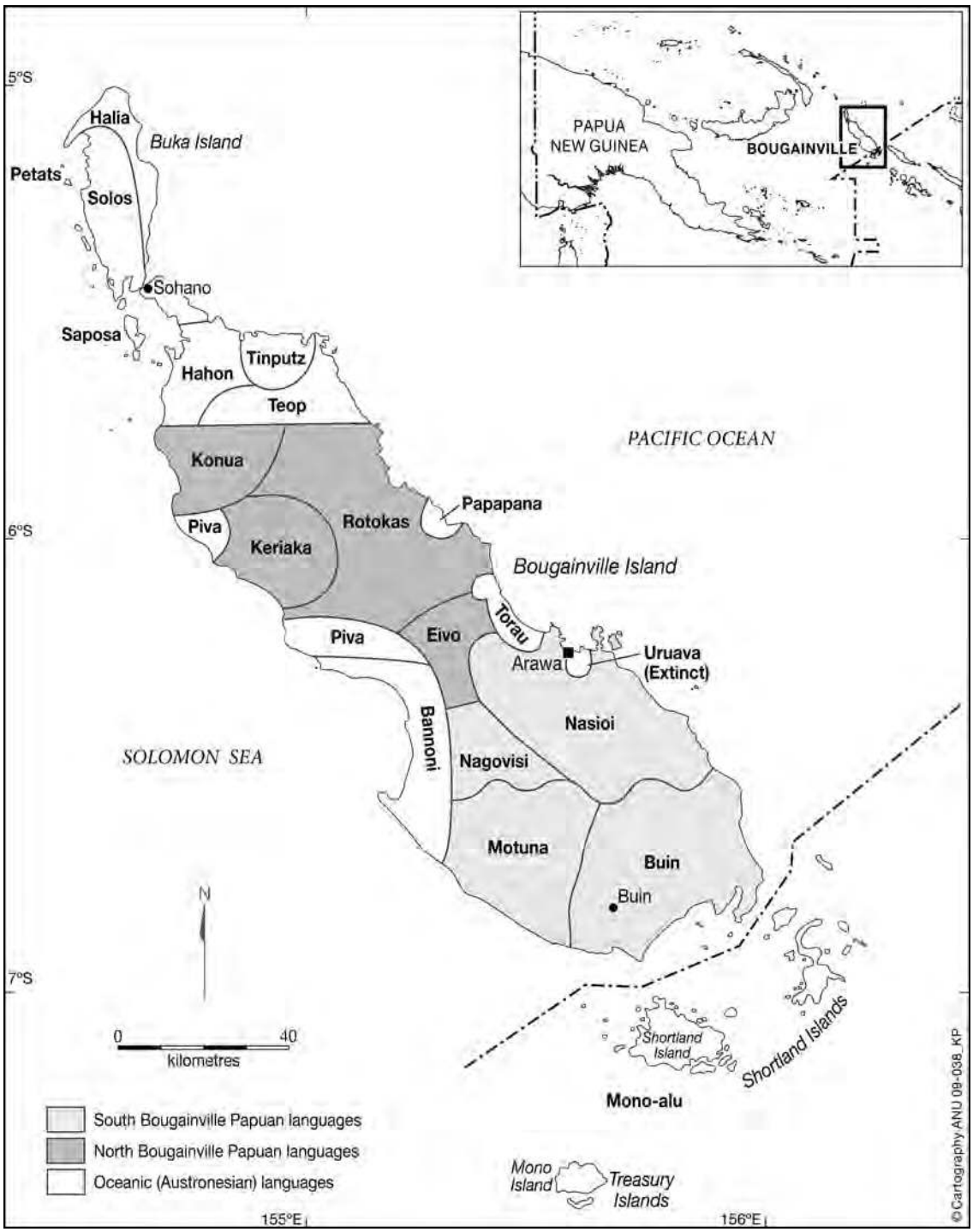
Ross' South Bougainville family comprises four languages, Nagovisi, Nasioi, Buin and Motuna (Siwai), which are spoken in the Kieta and Buin districts of southern Bougainville (see Map 1). The genetic relatedness of these four languages had previously been proposed by Allen and Hurd (1965) on the basis of a lexicostatistical study, and by Wurm (1975, 1982) on the basis of shared lexical and structural characteristics.[4] Wurm (1975, 1982) also proposed that South Bougainville (his East Bougainville) consisted of two subgroups: (a) Nasioi comprising the Nagovisi and Nasioi languages; and (b) Buin comprising the Buin and Motuna languages. The evidence for these two groupings appears to be differences in percentages of shared lexicon, as set out by Allen and Hurd (1965). Thus Nasioi and Nagovisi share over 50 percent of basic vocabulary, while Buin and Motuna show percentages in the mid-thirties. In contrast, percentages of shared basic vocabulary between Nagovisi/Nasioi and Buin/Motuna are between the high teens and high twenties (Wurm 1975:792).

2.1 The data for this study

The data presented in this paper come from a number of different published and unpublished sources. The most detailed and accessible lexical data available for the South Bougainville languages is Laycock and Onishi's (2003) Buin dictionary. Laycock distinguishes six Buin dialects; northern, northeastern, southern, eastern and western dialects, as well as the central dialect on which the dictionary is based. Additional Buin

[4] Bougainville island lies along a northwest-southeast axis, and so Nagovisi, Nasioi, Buin and Motuna can be described as located in the southeast of the island. I follow Ross in labeling the language family South Bougainville in contrast to the North Bougainville family. This is different from Wurm who labeled these groups East Bougainville and West Bougainville, respectively.

data comes from: (a) Grace and Kanoai (1955), which consists of over 600 lexical items over 300 grammatical constructions (noun phrases, possessive phrases, simple sentences) from Uisai (northeastern) Buin; (b) Hall (n.d.), an extensive list of lexical items and sentences in 26 languages of Bougainville and the northern Solomon Islands, including Uisai Buin; and (c) Griffin's (1974) grammar sketch of the language.



Map 1: Languages of Bougainville

For Motuna, Onishi (2002) presents extensive lexical data collected by Douglas Oliver in 1938–39, annotated with his own more recent checking with contemporary speakers. These data are supplemented by several other sources. Grace and Timpa (1955) give comparable data to that of Grace and Kanoai (1955) described above, while Pawley and Iobo (1971) present a shorter list of around 215 items of basic vocabulary.[5] Gasaway (1987b) gives 120 items of basic vocabulary from five dialects of Motuna, while Gasaway (1987c) gives an additional 136 items. Some additional Motuna lexical items are from Hall (n.d.). Onishi's (1994) description of Motuna provides the most detailed grammatical information available on any of the South Bougainville languages.

The Nasioi data are primarily from three word lists; Grace and Davika (1955), Pawley and Bintamoi (1971) and Pawley and Kauori (1971). Additional lexical data has been taken from a number of other sources on Nasioi grammar, namely Müller's (1949), Hurd and Hurd (1970), Hurd (1977) and Hurd (1992).

The Nagovisi data are from Pawley and Namoko's (1971) word list, as well as word lists in Gasaway (1987a, c), Decker's (1981) grammatical description, and Hunt's (1992) phonological sketch.

This plethora of word lists from southern Bougainville languages provides the lexical data necessary for applying the Comparative Method. However, there are also a number of problems with using data from such a range of sources. First, the orthographic conventions of each word list are not always clear, and there are uncertainties regarding the phonological shape of some items.[6] Second, the data likely consist of a combination of archaic and contemporary lexical items and so the data for each language cannot be taken to represent a single coherent stage of the language. Third, detailed morphological analysis has not been undertaken for each of the languages. While analyses given in the different grammatical sources have been consulted, it is still possible that certain morpheme boundaries within the lexical data have been misanalysed. For example, verbal forms occur in the word lists with a range of suffixes in each language. Only verbal roots are compared here, and suffixes are identified on the basis of grammatical descriptions of each language or recurrent endings within the lexical data.

From the various sources on South Bougainville languages, a set of comparative lexical data was put together that consists of 377 meanings for which I had data from more than one South Bougainville language and from a range of semantic domains; see Table 1. It is from these data that regular sound correspondences amongst the languages were determined and cognate sets established.

---

5 Andrew Pawley collected basic vocabulary for a number of languages of Papua New Guinea, including Motuna, Nasioi and Nagovisi of southern Bougainville, by requesting school teachers to ask their students to complete his 215 item word list. These are cited here as co-authored by Pawley and the student who completed the word list.

6 The data given in the Appendix follow the orthography of each source, with the exception that sequences of identical vowels are represented by a single vowel with a macron.

Table 1: Semantic domains and number of meanings compared

Semantic domain	Number of meanings compared
body parts	40
numerals	13
kin and people	26
physical environment	40
flora	31
fauna	25
property terms	46
motion and stance	27
bodily function	26
basic actions and events	52
material culture	51
Total	**377**

Table 2 shows the number of cognate sets for each semantic domain across different groups and pairings of South Bougainville languages. These forms, as described in §2.2, show predominately regular sound correspondences. Columns A–E represent the number of cognate sets that provide support for the South Bougainville language family, with cognate forms in all four, or at least three, of the contemporary languages. There are also a significant number of apparent cognates found in only two of the four languages, and these are shown in columns F–K. The supporting data for each of these cognate sets are given in the Appendix.

Table 2: Number of cognate forms across different sets of contemporary languages

Semantic domain	A	B	C	D	E	F	G	H	I	J	K	**Total**
body parts	6	0	2	6	3	3	0	2	0	1	4	**27**
numerals	2	0	1	0	0	1	0	0	1	0	2	**7**
kin and people	3	0	4	5	1	2	0	0	0	0	2	**17**
physical environment	5	1	2	2	1	3	0	5	3	0	8	**30**
flora	2	1	2	2	2	0	1	2	0	0	5	**19**
fauna	2	0	1	0	0	1	0	2	1	0	0	**9**
property terms	6	2	3	1	1	8	0	3	1	0	2	**27**
motion and stance	1	2	1	1	1	3	0	1	0	0	8	**18**
bodily function	5	0	1	0	3	2	0	0	1	1	6	**19**
basic actions and events	5	0	1	2	4	3	2	3	3	3	5	**31**
material culture	1	0	0	0	3	0	0	4	1	1	1	**11**
Total	**42**	**6**	**18**	**19**	**19**	**26**	**3**	**22**	**11**	**6**	**43**	**215**

Note: Columns A–K give the number of cognate items across different sets of languages:

A	Nagovisi, Nasioi, Buin, Motuna	G	Nagovisi, Buin
B	Nagovisi, Nasioi, Buin	H	Nagovisi, Motuna
C	Nagovisi, Nasioi, Motuna	I	Nasioi, Buin
D	Nagovisi, Buin, Motuna	J	Nasioi, Motuna
E	Nasioi, Buin, Motuna	K	Buin, Motuna
F	Nagovisi, Nasioi		

2.2 Sound correspondences and phonological reconstruction

Tables 3 and 4 set out the consonant and vowel correspondences across the four South Bougainville languages, as well as the reconstructed Proto South Bougainville phoneme for each correspondence set. While there are irregularites, as described below, there is a high degree of regularity across all four languages.

Table 3: South Bougainville consonant correspondences

	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna	Environment
1	*m	m	m	m	m	
2	*p	p	p	Ø, h	h	_ V_{ROUND}
3	*p	p	p	p (Ø), h	h	_ $V_{NON\text{-}ROUND}$
4	*b	w	b	p	p	
5	*n	n	n	r, n	n	
6	*r	r, l	r, l	r	r	V _ V
7	*r	r, l	d	n, r	d, r	# _
8	*d	(n)d	nt	t	t	
9	*t	t	t	r	t	V _ V_{BACK}
10	*t	t	t	t	t	# _ V_{BACK}
11	*t	s	s	t, s	s	_ i
12	*t	s	s	t, s	h	_ i
13	*ŋ	Ø	ŋ	Ø	Ø	_ #
14	*ŋ	ŋ	–[A]	ŋ	ŋ	
15	*k	k	k	k	k	
16	*g	(ŋ)g	–	g	ŋ	

[A] A dash in the correspondence sets indicates that no cognate forms have been found in the current sample of data.

The regular and unchanging correspondence of *m* across all four languages, and thus the reconstruction of Proto South Bougainville *m, is supported by nearly 40 cognate sets; see for example ‘coconut tree’ (Table 13) and ‘cold’ (Table 15).

The reconstruction of Proto South Bougainville *p, based on the two correspondences given in rows 2 and 3 of Table 3, is supported by only a handful of cognate sets. The correspondence of Nagovisi and Nasioi *p* with Buin *h* ~ *Ø* and Motuna *h* before a round vowel is supported by four cognate sets, including Nagovisi *pann*, Nasioi *paniŋ*, Buin *hō* ~ *ō* and Motuna *hoŋ* ‘sky’, reflecting Proto South Bougainville *pɔn(iŋ) (see Table 12).[7] The *h* ~ *Ø* variation in Buin appears to represent a dialectal difference, with forms showing the loss of *p found in the Central Buin data in Laycock and Onishi (2003), while the Uisai Buin data in Grace and Kanoai (1955) consistently show *h*. The correspondence set in row 3 is less well-supported, and further data is needed to establish the regular reflex of Proto South Bougainville *p preceding a non-round vowel in Buin. In these data, only a single etymon with *p in this phonological environment and a reflex in the Central Buin data has been found. This is the cognate set for ‘hit’ represented by Nasioi *nipo*, Buini

[7] It is unclear if the final *-iŋ* segments of the Nasioi form *paniŋ* ‘sky’ should be reconstructed as part of the protoform, and so it is included in parentheses in the reconstruction. The change from *n to *ŋ* in Motuna reflects the neutralisation of nasal contrasts as *ŋ* word-finally in this language (Onishi 1994:14–16).

nipi ~ nipikale and Motuna *nihkarei* (Table 18), indicating that in this environment *p is retained as such in Buin. However, additional data suggests that the reflexes of *p in Buin, as in the other three languages, may be identical in all phonological environments and that the retention of *p in *nipi ~ nipikale* 'hit' is irregular. Thus data from Grace and Kanoai (1955) indicates that the correspondence Nagovisi and Nasioi *p*, Uisai Buin *h*, and Motuna *h* also occurs before non-round vowels; see, for example, the cognate set for 'sand', including Nagovisi *pisima*, Nasioi *pīsi*, Uisai Buin *hisia* and Motuna *hīsia ~ hisia* (Table 12). Also there are a few cognate sets shared only by Buin and Motuna which show the expected Central Buin *Ø*, Uisai Buin *h* and Motuna *h* correspondence, including one in the environment of a following non-round vowel, namely Central Buin *ī*, Uisai Buin *hī* and Motuna *hi* 'hair' (Table 9).

A voiced bilabial stop *b is also reconstructed for Proto South Bougainville based on the correspondence of Nagovisi *w*, Nasioi *b* and Buin and Motuna *p*.[8] This correspondence is regular and is supported by more than ten cognate sets.[9]

Comparison of the alveolar stop, nasal, fricative and rhotic segments across the South Bougainville languages is complex, with the same segments occurring in two or more correspondence sets, as shown in rows 5–12 of Table 3. Some of this complexity is due to orthographic conventions. That is, certain allophonic realisations are represented orthographically. For example, in Nasioi the difference between *r, l* and *d* (rows 6 and 7) reflects different allophonic realisations of /d/, which tends to occur as [d] word-initially and following a glottal stop, as [ɾ] word-medially following a vowel, and as [l] word-initially before /u/ (Hunt 1992). Although the orthographic difference between *r* and *l* in the wordlists is recorded in the sound correspondences it will not be discussed further here given that a phonemic distinction between [r] or [ɾ] and [l] is not found in any of the four languages.[10] For the eight alveolar correspondence sets in Table 3, only four protophonemes are reconstructed, namely *n, *t, *d and *r.

Proto South Bougainville *n is reconstructed for the correspondence set in row 5 of Table 3; Nagovisi, Nasioi and Motuna *n* and Buin *r ~ n*. The *r ~ n* variation in Buin again reflects dialectal variation, noted by Laycock and Onishi (2003:xiii) as a difference between Central and Northern Buin dialects, respectively. In the data presented here those Buin forms taken from Laycock and Onishi's (2003) dictionary of Central Buin show *r*, and the Uisai Buin data from Grace and Kanoai (1955) show *n*. While fourteen cognate sets support the correspondence given in row 5 of Table 3, there are a couple of forms which are exceptional. Central Buin *nunno* 'liver' (Table 9) and *nipi* 'hit' (Table 18) both have an unexpected *n* word-initially. The geminate *nn* in *nunno* 'liver' is also unexplained.

Correspondence sets 6 and 7 in Table 3 appear to be in complimentary distribution and so a single protophoneme *r is reconstructed. The correspondence of *r* (realised as either

---

8 Gasaway (1987a:4) notes that there is dialectal variation between [b] and [w] for some lexemes. For example, she records Lamane Nagovisi *wūʔ*, To'mau Nagovisi *buʔ* and Border Nagovisi *wuʔ* for 'river'. With other lexemes the same pronunciation is found across all dialects and in yet other lexemes there is a phonological contrast between /b/ and /w/.

9 In one of the Motuna word lists, Pawley and Iobo (1971), two lexemes which form part of the cognate sets of this correspondence are written with the symbol *b*. In Motuna the voiced bilabial stop segment is the post-nasal allophone of /w/ (Onishi 1994:16).

10 As noted, in Nasioi [l] tends to occur word-initially before /u/. Gasaway (1987a:19) reports for Nagovisi that [l] tends to occur word-initially and [r] medially, but notes that both pronunciations could occur in both positions. For Buin it is reported that [l] and [ɾ] are in free variation (Laycock and Onishi 2003:xiii), while Onishi (1994:8, 18) records [l] only in loanwords in Motuna.

[r], [ɾ] [l]) across all four languages is supported by nearly 20 cognate sets, all of which comprise *r* intervocalically; see for example the cognates sets for 'head' (Table 9) 'eel' (Table 14) and 'to hear' (Table 17) This complements the correspondence Nagovisi *r* ~ *l*, Nasioi *d*, Buin *r* ~ *n* and Motuna *d* ~ *r* which occurs in word-initial position and is supported by only five cognate sets, including that for 'ear' (Table 9) 'sun' (Table 12) and 'to breathe, live' (Table 17).[11] These correspondences represent synchronic allophonic variation in the languages. As mentioned Nasioi /d/ is realised as [d] word-initially and as [ɾ] word-medially following a vowel, and it is this synchronic pattern which is represented in the sound correspondences in Table 3. Onishi (1994:17–18) describes how the /r/ phoneme in Motuna may be realised as [d] word-initially, but that it is in free variation with [ɾ] or [r] in this environment. This explains the occurrence of both *d* and *r* in Motuna in correspondence set 7; see for example, the cognate set for 'ear' (Table 9). The Buin data for correspondence set 7 is limited to four of the possible five cognate sets, and on the basis of these data it is unclear what factors condition the *r* ~ *n* variation. Both *r* and *n* are found in the Central and Uisai Buin data, and there is not necessarily agreement across the dialects for a particular etymon. For example, Central Buin has *rū* 'to fall' with an initial *r*, while Uisai Buin has *nukale* 'to fall' with an initial *n*.

Only a few cognate sets provide support for correspondence set 8 in Table 3 and only two of them have cognates in all four languages. Since the correspondence of Nagovisi (*n*)*d*, Nasioi *nt*, Buin and Motuna *t* contrasts with the other alveolar correspondence sets, a separate phoneme *d is reconstructed for Proto South Bougainville.

In the Nagovisi data the same lexemes are recorded with both prenasalised and oral voiced alveolar stops, such as *panda* ~ *pada* 'leaf (Table 13) Gasaway (1987a) compares wordlists from four areas within the Nagovisi-speaking region and suggests that the occurrence of prenasal and oral stops reflects dialectal differences. She notes (1987a:1–2) that a mountain dialect, Lamane, shows predominately prenasalised stops, while two dialects to the west, To'mau and Beretembe, show predominately oral stops, but in none of these varieties is the difference found to be contrastive. An eastern dialect, spoken at Sindalu, shows both prenaslised and oral voiced stops, and here Gasaway (1987a:2) suggests that there might be a phonemic contrast, but does not elaborate. Gasaway (1987a) in her brief description of Nagovisi prenasalised and oral stops gives examples of bilabial, alveolar and velar segments. However, in the present data, this variation is most apparent with alveolar stops, and without more detailed contemporary data the role and histories of prenasalised stops in these languages are not clear. Based on considerations of the most plausible kinds of sound change in the history of the contemporary languages and the system reconstructed for Proto South Bougainville, it seems best to reconstruct a series of voiced oral stops, namely *b, *d and *g. Proto South Bougainville *d is reflected in Buin and Motuna as *t*, having undergone the same devoicing change reconstructed for the bilabial stop series. These apparently parallel changes provide additional support for the reconstruction of *d and its change to *t* in Motuna and Buin, which is otherwise supported by only two cognate sets. The regular correspondence between Nagovisi (*n*)*d* and Nasioi *nt* is additionally supported by few other cognate sets found only in these two languages. The presence of the apparent nasal-stop sequence in Nasioi and of prenasalised representations in a number of the Nagovisi lexemes does suggest that in Proto South Bougainville *d, at

[11] In all five cases *r* occurs word-initially before a back vowel. It is thought that this is an artefact of the limited sample of data, but further research is needed to confirm the generalised environment of word-initial position for this correspondence set.

least, may have had both oral and prenasalised pronunciations. As the discussion indicates, the reconstruction of *d is less certain than that of other phonemes, which are supported by more tokens of the correspondence set.

For the final four alveolar correspondence sets, rows 9–12 in. Table 3, the single Proto South Bougainville phoneme *t is reconstructed. These four correspondence sets occur in complimentary environments. First, correspondence set 9 comprises *t* in Nagovisi, Nasioi and Motuna and *r* in Buin and, although illustrated by only three cognate sets, it always occurs intervocalically where the following vowel is a back vowel, that is *o*, *u* or *a*. This contrasts with correspondence 10, *t* in all four languages, which occurs word-initially before a back vowel. For example, Buin *turu(pai)* 'to spit' corresponds to Nasioi *tutu(pa)* and Motuna *tūtū(ha),*[12] such that word-initially there is a correspondence between Buin *t* and Nasioi and Motuna *t*, whereas intervocalically there is a correspondence between Buin *r* and Nasioi and Motuna *t* (see Table 17). However, it should be noted that there are also a number of apparent Buin cognates that show irregular retention of *t* intervocalically before a back vowel, such as *taita* 'brother (older of male)' (Table 11). Correspondence sets 11 and 12, Nagovisi and Nasioi *s*, Buin *t* ~ *s* and Motuna *s* or *h*, occur only in the environment of a following high front vowel *i*.[13] The variation in Buin again reflects dialectal variation; Central Buin forms show *t*, while Uisai Buin forms show *s*. Thus Central Buin *ti* 'to put, place' and Uisai Buin *sisale* 'to put' correspond to Nagovisi *sīgsi* 'to put' and Motuna *sirarei* 'to put' (Table 18). Correspondence set 11 is regular and supported by at least seven cognate sets. However, there are four other cognate sets that illustrate the correspondence given in row 12 of Table 3. In these cases, not only is the original consonant reflected as *h* in Motuna, the following vowel is also lost. Thus Central Buin *koti* 'garden; enclosure' and Nagovisi *kasiʔ* 'garden' correspond to Motuna *koh* 'garden' (Table 12). Conditioning environments which distinguish correspondence sets 11 and 12 are not apparent from the current sample of data.

The two correspondence sets in rows 13 and 14 of Table 3 are reconstructed as reflecting the single Proto South Bougainville phoneme *ŋ. The presence of a velar nasal in Nasioi and the lack of any corresponding segment in the other three languages occurs regularly in word-final position, and is supported by seven cognate sets, including 'blood' (Table 9), 'sea' (Table 12) and 'sugarcane' (Table 13). At this stage, I reconstruct word-final *ŋ in such contexts, despite its retention in only a single language and apparent loss in the other three languages. In the data, there are only two cognate sets, namely 'man' (Table 11) and 'black' (Table 15), which illustrate the correspondence of a velar nasal in Nagovisi, Buin and Motuna, and for neither has a Nasioi cognate been found. Thus additional data is needed before any claims can be made regarding the history of velar nasals in these languages.

There are two correspondence sets for velar stops, rows 15 and 16 in Table 3. The reconstruction of Proto South Bougainville *k for the regular correspondence of *k* across all four languages is supported by around twenty cognate sets; see for example 'four' (Table 10), 'garden' (Table 12), 'to be warm, hot' (Table 15) and 'to bite' (Table 18). Also reconstructed for Proto South Bougainville is a contrasting velar stop *g. The correspondence set supporting this reconstruction is that in row 16 of Table 3. However, it

[12] For 'to spit' in Motuna there is variation across the word lists. Although the form given above has a medial *t*, a medial *r* is found in a different word list.

[13] In the current data there are no cognate sets which illustrate the occurrence of the alveolar stop preceding the front mid vowel *e*.

is based on only two uncontroversial cognates sets, both of which lack Nasioi forms. In contemporary Nagovisi and Buin voiced and voiceless velar stops are described as separate phonemes, and this distinction is also reconstructed for Proto South Bougainville since correspondence sets 15 and 16 occur in contrastive environments. In Motuna, which has only a single series of stops at all places of articulation, there appears to have been a change from an oral stop *g to a nasal stop *ŋ*. For example, Motuna *koriŋi* 'fish (generic)' is cognate with Nagovisi *kare(**ŋ**)**g**e* 'fish' (Table 14) and Motuna *nuŋaŋ* 'male' is cognate with both Buin *ru**g**aŋ* 'man (PL)' and Nagovisi *nu**g**aŋo* 'man' (Table 11).

The vowel correspondences across South Bougainville languages are also very regular (see Table 4), although there are a number of complications relating to differences across the synchronic vowel systems and to the varying representations of forms across the different sources of data. First, all four languages appear to allow sequences of two vowels, including identical ones (Hurd 1992; Gasaway 1987a; Onishi 1994; Laycock and Onishi 2003). However, it is not clear that sequences of identical vowels have been consistently recorded in all the word lists. Thus, although it seems likely that Proto South Bougainville also had sequences of identical vowels, and there are a few cognate sets where a sequence of identical vowels is recorded in two or more languages, including 'knee' (Table 9), 'betelnut' (Table 13) and 'to flow' (Table 16), further and more accurate synchronic lexical analysis is needed.[14] Although sequences of non-identical vowels occur in etyma in a number of the cognate sets, in these data there are no recurring patterns of correspondence with a sequence of two non-identical vowels in one language corresponding to a sequence of identical vowels and/or a single vowel in the other languages.[15]

Contemporary Nasioi, Buin and Motuna have a typical five-vowel system with *i*, *e*, *o*, *u* and *a* (Hurd and Hurd 1970; Laycock and Onishi 2003; Onishi 1994). Gasaway (1987a) and Hunt (1992) present similar analyses of a five-vowel system for Nagovisi.[16] Decker (1981), on the other hand, suggests a six vowel system for Nagovisi, which she represents with the symbols *i*, *e*, *a*, *o*, *ö* and *u*. Decker (1981) does not provide a description of the realisations of each of the six vowels, and her data indicate that *ö* has a different distribution from the other vowels. That is, while the other five vowels occur in Decker's (1981) data singularly and in sequences of identical and non-identical vowels, *ö* occurs only in a sequence of two identical vowels. Decker's (1981) data does, however, contain a number of minimal or near minimal pairs, suggesting that *ö* is a distinct phoneme, including *moo* 'coconut', *möö* 'to get', *mo* 'I come', *muuga* 'at night' and *nnaa* 'my wife'.

---

[14] In Motuna all sequences of two vowels, including identical vowels, are best analysed as such, rather than as diphthongs or long vowels (Onishi 1994:18–22). In Nagovisi sequences of identical vowels in which the stress in on the second vowel are analysed as a sequence of two individual vowels, while sequences which are unstressed or in which stress is on the first vowel are analysed as monosyllabic long vowels (Gasaway 1987a). As stress is not indicated in any of the sources of data used here, all Nagovisi items are represented as a sequence of two identical vowels and not as a single long vowel. The same representation is used for the other three languages, as well as the reconstructed forms.

[15] For the Proto South Bougainville, a sequence of non-identical vowels is reconstructed if two or more languages show such a sequence. The reconstructions thus posit an hypothesis that is economical in terms of number of changes proposed to have occurred, but requires further investigation of recurring patterns of correspondence across the contemporary languages.

[16] Note that Hunt (1992) describes the Nagosivi vowel system as comprising *i*, *e*, *a*, *ɔ* and *o*, while Gasaway (1987a) indicates that Nagovisi has the same five-vowel system as the other three languages.

Table 4: South Bougainville vowel correspondences

	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
1	*u	u	u	u	u
2	*o	o	o	u	u
3	*o	o	o	o	o
4	*ɔ	a	a	o	o
5	*a	a	a	a	a
6	*i	i	i	i	i
7	*e	e	e	e, i	i

The patterning and occurrence of the seven correspondence sets, given in Table 4, suggest the reconstruction of six vowel phonemes for Proto South Bougainville. Correspondence sets 1–5 show different patterns of correspondence amongst the three back vowels found in the contemporary languages, namely *u*, *o* and *a*. Correspondence set 3 is represented by four cognate sets only, and will be discussed after the other four correspondence sets have been described.

Correspondence sets 1, 2 and 4 in Table 4 are supported by around 20 cognate sets each, while correspondence set 5 is supported by 34 cognates. Each of these sound correspondences occurs in a wide range of phonological environments taking into account position in the word (e.g. initially, medially, finally) and surrounding consonant and vowel segments. While there are no exact minimal pairs across these four correspondences, the presence of near minimal pairs does indicate that these four sound correspondences are contrastive and so represent distinct protophonemes. This is illustrated by the cognate sets in Table 5. For example, the cognate sets for ‘head’, ‘bird’ and ‘menstrual house’ demonstrate the occurrence of correspondence sets 2, 4 and 5 in the initial syllable of a word, preceded by a labial consonant, followed by rhotic consonant and with a front vowel in the following syllable. The reconstruction of Proto South Bougainville back vowels, reflected by these four correspondence sets, appears to be reasonably straightforward. Correspondence sets 1 and 5 comprise the same vowel segment across all four languages and these two segments, *u and *a respectively, are reconstructed as the Proto South Bougainville phonemes. The correspondence of Nagovisi and Nasioi *o* and Buin and Motuna *u* (set 2), is taken to reflect an original *o segment, while the correspondence of Nagovisi and Nasioi *a* and Buin and Motuna *o* (set 4) is taken to reflect an original *ɔ. The contrast of correspondences 2 and 4 indicates that two back vowel phonemes can be reconstructed for Proto South Bougainville alongside *u and *a, and reconstructing *o and *ɔ posits more plausible changes to account for the contemporary vowel forms than other reconstructions.

Table 5: Cognates illustrating contrastive environments of back vowel correspondences[17]

Gloss	Nagovisi	Nasioi	Buin	Motuna	Corr. set
eye	**u**ta	d**u**ta	—	**u**to	1
child	t**ō**tō	t**o**-ʔto	—	t**u**tu	2
island	m**a**tomato	—	m**o**ruroi	m**o**tuka(h)	4
hear	t**a**la(nsi)	t**a**ra	—	t**a**rowarei	5
tobacco	—	b**u**rusi	p**u**rutū	p**u**ruweku	1
daughter	(n)**o**ra	—	r**u**ro	n**u**ro	2
big	p**a**nna	p**a**nkaiŋ	**o**ŋokopa	h**o**nno	4
eel	w**a**rama	b**a**ra-ʔma	p**a**lamo	p**a**ramo	5
old (thing)	**u**likala	**u**rika	**u**ni	—	1
fear	—	**a**rorō	**o**uru	**o**ruharei	4
sleep	**a**si	**a**si	**ā**ti	**a**tihe	5
push	t**u**mele	t**ū**mē	—	t**ū**miwarei	1
ear	l**o**m, r**o**ŋ	d**o**me	n**u**me	d**u**ŋ	2
left (hand)	—	m**a**reʔānāŋ	m**o**ripere	m**o**rinonna	4
cold	kam**a**li	kam**a**ri	kam**a**li	kam**a**ʔ	5
head	w**o**re	b**o**re	p**u**re	p**u**ri	2
bird	w**a**re(ŋ)ge	b**a**reŋ	—	p**o**riŋi	4
house (menstrual)	—	b**a**re	p**a**le	p**a**ri	5
breathe, live	l**o**masiaisi	d**o**mantū	n**u**matu	d**u**maruharei	2
dust	l**a**mo	d**a**moŋ	—	d**o**mu	4
eel	war**a**ma	bar**a**-ʔma	pal**a**mo	par**a**mo	5

Correspondence set 3 is problematic. It is supported by four cognate sets and comprises the correspondence of the back *o* vowel across all four languages. However, none of the cognate sets are complete or regular; see Table 8. For example, while in the cognate set for 'to know, think', including Nasioi, Buin and Motuna forms, there is a correspondence of *o* in these three languages, in the other three cognate sets, which include Nagovisi forms with the vowel *o*, the Nasioi and Buin forms comprise vowel sequences.[18] These correspondences are analysed as reflecting Proto South Bougainville *o, although further data are needed to ascertain if there are different conditioning environments for sound correspondence sets 2 and 3, if there are recurrent correspondences of vowel sequences across the four languages, as well as if the cognate sets in Table 8 do represent regular patterns of sound change or irregularities.

Table 6: Cognates illustrating the back vowel correspondence set 3[19]

Gloss	Nagovisi	Nasioi	Buin	Motuna
know, think	—	**o**noukō, **o**nouʔ	**o**no, **o**nohale	**o**noharei, **o**nohihe
betelnut	m**ō**sī	m**oi**si	—	m**o**si, m**ō**si
guts	k**ō**	—	k**ou**	k**o**
shoot	t**ō**ʔ	t**ō**(kō)	t**ua**, s**ua**kale	t**ō**hēwarei, t**o**hihe

[17] The complete supporting data for each of these cognate sets is given in the Appendix. For reasons of space only one cognate form from each language is given here.

[18] The cognate set for 'breast' (Table 9) also appears to support this correspondence, but cognates have been found only in Nagovisi and Motuna.

[19] Explanation of the variant forms for each language is provided with the tables of supporting data in the Appendix.

There are only two vowel correspondence sets comprising the front vowels *i* and *e*; rows 6 and 7 in Table 4. They support the reconstruction of the two phonemes *i and *e for Proto South Bougainville. The reconstruction of *i is straightforward, with 24 cognate sets illustrating the regular correspondence of *i* in the four languages. The correspondence of Nagovisi, Nasioi and Buin *e* and Motuna *i* is less well-supported, but is found in nine cognate sets. That these two correspondence sets are contrastive and so reflect two original protophonemes can be demonstrated by near minimal pairs. For example, the cognate set that includes Nagovisi *wore*, Nasioi *bore*, Buin *pure* and Motuna *puri* 'head' illustrates the correspondence of *i* in Motuna and *e* in the other three languages, while the cognate set that includes Nagovisi *(n)uri*, Buin *nuni* and Motuna *nuri* 'son (my)' shows the correspondence of *i* across the languages. There are also three cognate sets in which there is a correspondence between Nagovisi and Nasioi *e* and Buin and Motuna *i*. Comparison of the cognate set for 'head' (given above) with that for 'left (hand)' which includes Nasioi *mar**e**ʔānāŋ*, Buin *mor**i**pere* and Motuna *mor**i**nonna*, shows that the occurrence of *e* or *i* in Buin does not appear to be determined by the phonological environment of the segment. These three cognate sets are analysed here as representing irregular change of Proto South Bougainville *e to Buin *i*. However, analysis of additional data is needed to ascertain if they are representative of a third correspondence set of front vowels.

These established sound correspondences and their analysis as reflexes of a set of Proto South Bougainville phonemes appear to provide a reasonable phonological history of the contemporary languages. However, some contemporary phonemes are not accounted for, including the glottal stop in Nagovisi, Nasioi and Motuna and the Motuna vowel *e*. Nagovisi and Nasioi have a phonemic glottal stop (Hunt 1992; Hurd 1992), while in Motuna the glottal stop has both phonemic and phonetic realisations (Onishi 1994:12–13). Central Buin, on the other hand, does not have a glottal stop phoneme (see Laycock and Onishi 2003:xiii). Glottal stops do not appear to have been consistently recorded in all the sources of data used here, and so it has not been possible to include this segment in the sound correspondences. More accurate contemporary data are needed before the history of this segment can be determined. As mentioned Motuna has a five vowel system of *i*, *e*, *o*, *u* and *a*. However, the historical analysis presented here accounts only for four of the Motuna vowels, namely *i, o, u* and *a*, but not *e*. On the basis of these data the history of Motuna *e* is unclear and is in need of further investigation.

2.3 Proto South Bougainville and sound changes in the contemporary languages

The discussion of sound correspondences across the South Bougainville languages confirms Ross' (2001, 2005) hypothesis that these languages are genetically related. In addition, this lexical study allows for the reconstruction of the Proto South Bougainville phoneme system and of certain sound changes that have occurred in the history of the contemporary languages.

The phoneme system reconstructed for Proto South Bougainville is given in Table 7. The consonant inventory is very similar to that of all four contemporary languages, although the voicing distinction in oral stops has been lost in Motuna and retained only for the velar stop in Buin. As mentioned, it is possible that Proto South Bougainville also had a glottal stop phoneme, as found in Nagovisi, Nasioi and Motuna, but this is not certain from the present data. Six vowel phonemes appear to be reconstructable for Proto South Bougainville, represented here as the five cardinal vowels present in the contemporary languages and an additional back mid vowel.

Table 7: Proto South Bougainville phoneme inventory

Consonants			Vowels	
*p	*t	*k	*i	*u
*b	*d	*g	*e	*o
*m	*n	*ŋ		*ɔ
	*r		*a	

The reconstruction of such a set of protophonemes implies that certain sound changes have occurred in the daughter languages. These are set out in Table 8. In all four languages there is a change from *t to *s* before the high front vowel *i*, and it seems likely that this represents allophonic variation that was present in the protolanguage.

Table 8: Sound changes in South Bougainville languages

Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
*p	—	—	*p > Ø (Central B.) *p > h (Uisai B.)	*p > h
*b	*b > w	—	*b > p	*b > p
*t	*t > s /_i	*t > s /_i	* t > r /V_V_{BACK} *t > s /_i (Uisai B.)	*t > s /_i (*ti > h)
*d	—	*d > nt	*d > t	*d > t
*k	—	—	—	—
*g	—		—	*g > ŋ
*m	—	—	—	—
*n	—	—	*n > r (Central B.)	—
*ŋ	*ŋ > Ø /_#	—	*ŋ > Ø /_#	*ŋ > Ø /_#
*r	—	*r > d /#_	*r > r, n /#_	*r > d, r /#_
*u	—	—	—	—
*o	—	—	*o > u	*o > u
*ɔ	*ɔ > a	*ɔ > a	*ɔ > o	*ɔ > o
*a	—	—	—	—
*i	—	—	—	—
*e	—	—	(*e > i)	*e > i

2.4 Internal relationships of South Bougainville

Wurm (1975, 1982) proposes that the South Bougainville family consists of two subgroups: Nasioi, comprising the Nagovisi and Nasioi languages, and Buin, comprising the Buin and Motuna languages.[20] His evidence for these groupings, based on Hurd and Allen (1965), is that Nagovisi and Nasioi share a higher percentage of basic vocabulary with each other than they do with either Buin or Motuna, and that the same is true of Buin and Motuna. Table 2 sets out the number of shared cognates across different groups of South Bougainville languages in the current sample of data. Of the 215 cognate sets, 104 occur in three or four languages and are taken to reflect original Proto South Bougainville forms. Of the remaining 111 cognate sets, contemporary forms are found in only two of the four languages. Buin and Motuna do share a significant number of cognate forms that are not found in either Nasioi or Nagovisi; 43 cognate sets as shown in column K of Table 2. However, the number of cognate sets exclusive to Nasioi and Nagovisi (26, column F in Table 2), is not significantly higher than the number shared by Nagovisi and Motuna (22, column H in Table 2). This raises questions about the subgrouping of Nasioi and Nagovisi.

There is also a methodological concern with using exclusively shared lexical items to define subgroups, particularly within a family of only four languages. The strongest evidence for subgrouping comes from exclusively shared innovations, which are best accounted for as occurring in a common ancestor distinct from that of languages outside of the proposed subgroup. While it is possible, and indeed likely, that at least some of the lexical cognates shared exclusively by Buin and Motuna are shared innovations, this is difficult to demonstrate. With only four languages in the family, and possibly two primary subgroups, it is nearly impossible to distinguish between shared innovations and shared retentions and thus use lexicon as diagnostic of subgrouping patterns. For example, if Buin and Motuna show a pair of cognate forms (e.g. *hī* and *hi* 'hair') and Nagovisi and Nasioi show a different pair of cognate forms for the same lexical meaning (e.g. *lapo* and *dapo* ~ *rapo* 'hair'), then without presuming a particular subgrouping it is impossible to determine which cognate set is retained from Proto South Bougainville, if any, and which reflects a shared innovation. If, on the basis of number of exclusively shared lexical items, Buin and Motuna are taken to form a subgroup, while Nagovisi and Nasioi are not, then it could be argued that the Buin and Motuna cognates represent a shared innovation in contrast to the retained forms in Nagovisi and Nasioi. However, if Nagovisi and Nasioi are also taken to form a subgroup, then it is impossible to determine which subgroup has retained the Proto South Bougainville form and which has the innovative form.

The reconstruction of the Proto South Bougainville phoneme system presented here does support a Buin-Motuna subgroup. As can be seen from Table 8, Buin and Motuna share a number of phonological innovations distinct from either Nasioi or Nagovisi:

(a) Proto South Bougainville *p has become h, with subsequent loss in Central Buin;

(b) Proto South Bougainville *b and *d have been devoiced, and *d merges with original *t;

(c) Proto South Bougainville *o has become u, merging with original *u; and

[20] Wurm (1975, 1982) uses different terms for groupings of genetically related languages based on the percentage of shared lexicon. Thus he labels South Bougainville (his East Bougainville) a 'stock', with between 12–20% shared vocabulary and the Nasioi and Buin groups 'families', with between 20–28% shared vocabulary. I use the term 'family' to denote a group of genetically related languages that have not been demonstrated to be related to any other languages, and the term 'subgroup' for a group of languages within a family that are more closely related to each other than to any other language within the family.

(d) Proto South Bougainville *ɔ has become o.

Innovation (a), *p > *h*, must have occurred before innovation (b) since the devoicing of original *b has not resulted in a merger of Proto South Bougainville *b and *p in these two languages. Similarly, innovation (c) must have occurred before innovation (d), as these two protophonemes have not merged in the contemporary languages.

Evidence for a Nasioi-Nagovisi subgroup is much weaker, supported here by a single phonological innovation, namely the change of Proto South Bougainville *ɔ to Nasioi and Nagovisi *a*. It should also be noted that there is an apparent sound change, loss of *ŋ in word-final position, which is shared by Nagovisi, Buin and Motuna.

In summary, the phonological evidence supports the hypothesis that Buin and Motuna form a subgroup within South Bougainville, but only weakly supports a Nasioi-Nagovisi subgroup and raises questions regarding a possible relationship between Buin and Motuna, and Nagovisi.

3 Concluding remarks

The lexical data examined in this paper confirm Ross' (2001, 2005) hypothesis that Nagovisi, Nasioi, Buin and Motuna of southern Bougainville form a language family, South Bougainville. This more detailed investigation of the family also allows for its internal relationships to be explored. There is strong phonological evidence that Buin and Motuna form a subgroup, while the evidence for a Nagovisi-Nasioi subgroup is much weaker and warrants further investigation.

The data presented here also raise questions regarding the role of sociocultural contact amongst southern Bougainville speech communities in the region's linguistic history. A number of lexical items in the South Bougainville languages appear to be borrowings from an Oceanic language. For example, Buin *mūre* and Motuna *mūri* 'back' (cf. Proto Oceanic *muri- 'back'), Nagovisi *polo*, Nasioi *poro*, Buini *ūru*, and Motuna *hūru* 'pig' (cf. Proto Oceanic *boʀok 'pig'), Nagovisi *susu* 'to suck' (cf. Proto Oceanic *susu 'to suck, suckle'), and Buin *īana* 'fish' (cf. Proto Oceanic *ikan 'fish'). While these forms indicate that contact with speakers of Oceanic languages played a role in the linguistic history of the South Bougainville languages, further research is needed to determine the nature and chronology of the contact or contacts. Some 'Oceanic' forms, like those for 'pig', occur in all four South Bougainville languages, follow the established sound correspondences, and do no appear to be direct borrowings from neighbouring Oceanic languages (cf. Torau *bo* 'pig' and Mono-Alu *boʔo* 'pig'; Palmer 2004, Tryon and Hackman 1983), and so appear to indicate ancient South Bougainville-Oceanic contact. Other forms appear to reflect more recent borrowings between currently neighbouring languages, such as Buin *īana* 'fish' and Mon-Alu *iana* 'fish' (Tryon and Hackman 1983). It is not just contact between Oceanic and South Bougainville speakers that has occurred. The figures of shared lexicon in Table 2 point to contact amongst speakers of the four South Bougainville languages. For example, while similar numbers of lexical items are exclusively shared by Nagovisi and its neighbouring languages Nasioi and Motuna, considerably fewer are shared by Nagovisi and Buin. It is likely that some of the lexical items that are exclusive to two neighbouring languages reflect borrowing, and thus contact, between adjacent speech communities. Distinguishing between retentions, shared innovations and borrowings in this context is not easy, but these matters need to be investigated in more detail in order to reconstruct an accurate linguistic history of South Bougainville.

References

Allen, Jerry and Conrad Hurd. 1965. Languages of the Bougainville district. Unpublished manuscript. Port Moresby: Department of Information and Extension Services.

Decker, Margaret. 1981. Notes on Nagovis grammar. Unpublished manuscript. Ukarumpa: SIL.

Dunn, Michael, Robert Foley, Stephen Levinson, Ger Reesink and Angela Terrill. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46(2).388–403.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink and Angela Terrill. 2008. Structural phylogeny in historical linguistics: methodological explorations applied in Island Melanesia. *Language* 84(4).710–759.

Gasaway, Eileen. 1987a. Notes on Nagowisi phonology and orthography. Unpublished manuscript. Ukarumpa: SIL.

—— 1987b. Notes on Siwai (Motuna) phonology and orthography. Unpublished manuscript. Ukarumpa: SIL.

—— 1987c. Comparison of Nagowisi and Siwai. Unpublished manuscript. Ukarumpa: SIL.

Grace, George and J. Davika. 1955. Linguistic questionnaire: Kieta. Unpublished manuscript. Honolulu: University of Hawai'i.

Grace, George and Luke Kanoai. 1955. Linguistic questionnaire: Uisai. Unpublished manuscript. Honolulu: University of Hawai'i.

Grace, George and Jeremiah Timpa. 1955. Linguistic questionnaire: Siuwai. Unpublished manuscript. Honolulu: University of Hawai'i.

Griffin, Margie. 1974. Notes on the Buin language. Unpublished manuscript. Ukarumpa: SIL.

Hall, A. n.d. Northern Solomons languages. Unpublished manuscript.

Hunt, Bruce, ed. 1992. Organised phonology data: Nagovisi language. Unpublished manuscript. Ukarumpa: SIL. [Available online: http://www.sil.org/pacific/png/pubs/0000262/Nagovisi.pdf]

Hurd, Conrad. 1977. Nasioi projectives. *Oceanic Linguistics* 16(2):111–178.

Hurd, Conrad, ed. 1992. Organised phonology data: Nasioi. Unpublished manuscript. Ukarumpa: SIL.

[Available online: http://www.sil.org/pacific/png/pubs/0000268/Nasioi.pdf]

Hurd, Conrad and Phyllis Hurd. 1970. Nasioi verbs. *Oceanic Linguistics* 9(1):37–78.

Laycock, Donald C. (ed. by Masayuki Onishi) 2003. *A dictionary of Buin. A language of Bougainville*. Canberra: Pacific Linguistics.

Müller, A. 1949. Nasioi grammar. Unpublished manuscript.

Onishi, Masayuki. 1994. A grammar of Motuna (Bougainville, PNG). PhD thesis. Canberra: The Australian National University.

—— 2002. *An annotated word list of the Motuna language (Bougainville, Papua New Guinea*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

Palmer, Bill. 2004. Torau draft dictionary. Unpublished manuscript. Surrey: Surrey Morphology Group. [Available online: http://www.surrey.ac.uk/lcts/bill.palmer/NWS_site/Tor/dic.htm#The%20dictionary]

Pawley, Andrew and Andrew Bintamoi. 1971. Language questionnaire: Nasioi. Unpublished manuscript. Auckland: University of Auckland.

Pawley, Andrew and William Iobo. 1971. Language questionnaire: Siwai. Unpublished manuscript. Auckland: University of Auckland.

Pawley, Andrew and John Kauori. 1971. Language questionnaire: Nasioi. Unpublished manuscript. Auckland: University of Auckland.

Pawley, Andrew and Margaret Namoko. 1971. Language questionnaire: Nagovis. Unpublished manuscript. Auckland: University of Auckland.

Ross, Malcolm. 2001. Is there an East Papuan phylum? Evidence from pronouns. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

—— 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew Pawley, Robert Attenborough, Jack Golson and Robin Hide, eds *Papuan Pasts. Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Pacific Linguistics.

Ross, Malcolm and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46(2).456–498.

Spriggs, Matthew. 1997. *The Island Melanesians*. London: Blackwell.

Summerhayes, Glen R. 2007. Island Melanesian pasts: a view from archaeology. In Jonathan Scott Friedlaender, ed. *Genes, language and culture in the southwest Pacific*, 1–36. New York: Oxford University Press.

Terrill, Angela. 2002. Systems of nominal classification in East Papuan languages. *Oceanic Linguistics* 41(1).63–88.

Tryon, Darrell T. and Brian D. Hackman. 1983. *Solomon Islands languages: an internal classification*. Canberra: Pacific Linguistics.

Wurm, Stephen A. 1975. The East Papuan Phylum in general. In Stephen A. Wurm, ed. *Papuan languages and the New Guinea linguistic scene*, 783–803. Canberra: Pacific Linguistics.

—— 1982. *Papuan languages of Oceania*. (Ars linguistica 7.) Tübingen: Gunter Narr.

Appendix: Cognate sets and Proto South Bougainville reconstructions

The following tables give the supporting data for all the cognate sets found in the data sample. If cognates occur in all four or in three of the four languages, this is taken as evidence for its presence in Proto South Bougainville and a lexical form is reconstructed. Cognates sets consisting of forms in only two of the contemporary languages are also included, but without additional data are not considered to be reconstructable.[21]

[21] If a lexical item occurs in only one or two of the available sources of data for a language, then it is indicated in the following tables by superscript letters after the lexical item. For Nagovisi: A - Decker (1981); B - Hunt (1992); C - Pawley and Namoko (1971); D - Gasaway (1987a, c); and E - Hall (n.d.). For Nasioi: F -

Table 9: South Bougainville cognate sets: body part terms

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
arm (hand)		uga[E, 1]	—	aku	aŋu
blood	*ereŋ	ele	ereŋ	irira, iriroi[J, 2]	iri
bone	*kōna	kōnāʔ, kona	koʔ[I], kō[G]	kāna[3]	kōna[M], konaʔ[P]
ear	*rome	lom[B], roŋ[B, D]	dome[F, G, I]	nume[J]	duŋ[M, N], ruŋ[M, O, P]
eye	*rutɔ	uta[4]	duta, ruta[I]	—	uto[4]
fat, grease	*titi	sisi	—	titi[J]	siʔ[N, P]
guts	*kō	kō[C]	—	kou	ko[M, O]
hand (arm)	*komɔ	ko(u)ma, kōma[D, 5]	—	kūmo[K, 6]	kumoputo[M], kumo[P,7]
head	*bore	wore	bore	pure	puri
knee	*mī	mī[D]	—	mī[J, 8]	mi[M], mī[P]
left (hand)	*mɔre-	—	mareʔānāŋ	moripere[J, 9]	morinonna
liver	*nonɔŋ	—	nonaŋ	nunno[J]	nuno[N, O]
neck	*kuru	kulu[C]	kuru	—	kuʔ[M, N], kū[O, P, 10]
nose	*keni	kenn[C], keŋ[D]	keni[G, I]	keŋ[J], keni[K]	kiŋ
right (hand)	*mē-	mēʔna[B]	—	maimaku[J, 11]	mēnonna
tongue	*meneŋ	mene[D]	meneŋ	mere[J]	mini
wing	*kupɔ	—	kupa	kupo[J], kupa[J], kuvo[K]	kūwo[N], kuwo[O]
back		vilö[C]	biʔloʔ[I]	—	—
flesh		wio[C]	bioʔ[I]	—	—
hair		lapo[C]	dapo[G], rapo[I]	—	—
belly		—	—	ere	ele
feather		—	—	paru	paru[12]
hair		—	—	ī[J], hī[K]	hi
mouth		—	—	muŋ[J]	muŋ
breast		nōnō[D]	—	—	nuno[M, P]
foot		tamma[C]	—	—	tompa[M, N]
egg		—	sirāŋ[G], siraʔŋ[I]	—	sira

1 Nagovisi *uqa* is given for the meaning ‘hand’ in Hall (n.d.).

2 Buin *irira* ‘blood’ is a Western dialect form, while *iriroi* is the term for ‘coagulated blood’ (Laycock and Onishi 2003).

3 Irregular change from *o > *a* in first syllable.

4 Irregular loss of initial consonant in Nagovisi and Motuna.

Müller (1949); G - Pawley and Kauori (1971); and I - Grace and Davika (1955). For Buin: J - Laycock and Onishi (2003); K - Grace and Kanoai (1955); and L - Hall (n.d.). For Motuna: M - Onishi (2002); N - Grace and Timpa (1955); O - Pawley and Iobo (1971); P - Gasaway (1987b, c); and Q - Hall (n.d.).

[5] Nagovisi *ko(u)ma* appears to have both the meanings of ‘hand’ and ‘palm’ (Gasaway 1987a). The variation in the vowel is dialectal.
[6] Buin *kūmo* means ‘hand’ (Grace and Kanoai 1955).
[7] Motuna *kumoputo* means ‘wrist’ (Onishi 2002), while *kumo* is listed as ‘hand, palm’ in Gasaway (1987b).
[8] Buin *mī* has the general meaning of ‘joint’ and can denote ‘knee’ and ‘elbow’ (Laycock and Onishi 2003).
[9] Buin *moripere* means ‘left side’, and *pere* is ‘side’ (Laycock and Onishi 2003).
[10] Motuna *kuʔ* and *kū* appear to have irregularly lost the final syllable of Proto South Bougainville *kuru ‘neck’.
[11] Buin *maimaku* is glossed as ‘right hand’ and contrasted with *maipere* ‘right side’ (Laycock and Onishi 2003).
[12] Also ‘flower’ in Motuna (Onishi 2002).

Table 10: South Bougainville cognate sets: numerals

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
three	*be-	wekago(ʔ)[1]	benaumo	paigami[J]	pekaŋ
four	*kɔre-	karekago(ʔ)	karenaumo	korigami	korikaŋ
ten	*nɔraŋ	nöra[a], nola[C, E]	naruŋ kivora	—	naraŋ
five		pa(ʔ)noko(ʔ)	panoko	—	—
one		—	—	nori[K, L]	no(ʔ)ri[M, 2]
hundred		—	—	pore	pore
thousand		—	naruŋ kokorei	kukurei	—

[1] The Nagovisi numerals ‘three’, ‘four’ and ‘five’ are represented with the glottal stops in Decker (1981) and without them in Pawley and Nakamo (1971) and Hall (n.d.).
[2] Irregular change from *a > *u* in the second syllable of Nasioi *naruŋ*.
[3] Motuna *no(ʔ)ri* denotes ‘noe’ for coconuts and appears to also refer to ‘one’ of other kinds of payments (Onishi 2002:268–269).

Table 11: South Bougainville cognate sets: kin and person terms

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
brother (older of male)	*batato	watata[C, E]	batāāta[G]	pararo (his)[L]	parato[N]
brother (older of male)	*tāta	—	tāta[I]	taita(nu) (my)[J]	tātā[M, 1]
child	*tōtō	tōtō[D]	to-ʔto[I]	—	tutu[N], tūtū[P]
daughter (my)	*norɔ	(n)ora	—	ruro[J], nulo[K]	nura[P], nuro[Q]
father	*bomɔ	woma (his)[E]	bauma[G]	pumo (his)[L]	pūmo[N], pumo (his)[Q]
husband	*bɔ[m,ŋ]	wamm[C]	bauŋ[G, H]	—	poŋ (her)
man	*nugaŋ	nugaŋo[D]	—	rugaŋ[J, 2]	nuŋaŋ
mother	*bōko	woko[C]	bauko[I]	—	pōku[N]
person	*nɔmm[e,ai]	namme[C]	—	—	nommai[O]
name	*mīŋ	mī[C]	—	miŋ[J]	mī
son (my)	*nuri	(n)uri	—	ruŋ[J], nuni[L]	nuri
wife	*bana	wana[C]	bāŋ[G, H]	para (his)[J]	pana (his)
brother (younger of male)		walamann[C]	baramanuŋ	—	—
name		miri[D]	miriŋ	—	—
chief		—	—	mumira[J], mumina[L]	mumih[Q]

[1] Motuna *tātā* refers to ‘elder brother’ (Onishi 2002).

[2] In Buin *rugaŋ* ‘man’ is the singular form, contrasting with *ruka* ‘men’ the plural form (Laycock and Onishi 2003).

Table 12: South Bougainville cognate sets: physical environment

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
cloud	*kɔmo	kamo	kamo	komui[K]	hom[M], hoŋ[M], komu[P]
dust	*rɔmo	lamo[C]	damoŋ[G, I]	—	domu[N, O]
garden	*kɔti	kasiʔ[D]	—	koti[J]	koh[N, O]
island	*mɔto	matomato[D]	—	moruroi	motuka(h)
ocean	*maira	—	maiara[I]	maiara[J]	maiara[M, N]
sand	*piti(a)	pisima[C]	pīsi[G, H], pisia[I]	hisia[K]	hīsia[M], hisia[N, O]
sea	*piruŋ	pilu[C]	piruŋ	—	hiru
sky	*pɔn(iŋ)	pann[C]	paniŋ	ō[J], hō[K]	hoŋ
smoke	*ī	ī	ī[I]	īto, īta[J, 1]	ī̄
sun	*rua	lō[D]	duā[G, H], ruah[I]	rua	—
water	*doŋ	ndo[C, D]	ntoŋ	tū[J], tu[K]	tu
ashes		sipu[C]	sipuŋ[H, I]	—	—
bush		pora(ʔ)	pora-a[H]	—	—
rain		apoʔo[C]	apo[H, I]	—	—
ashes		—	—	peu[J]	peu[M, O]
bush		—	—	muti(ne)	muhni
cave		—	—	nuinui[J]	nui=nui[M]
high tide, flood		—	—	rūkoro	rūworo[2]
mountain		—	—	menu	menu
river		—	—	tū[J]	tu[O]
road		—	—	monare[J]	monare[M, N], manare[O]
star		—	—	kaipa[J], kaiva[K]	kēwa
fog		maua[C]	—	—	muh[O]
grasslands		sikone[D]	—	—	sikani[P]
ground		mesi	—	—	misi
moon		pe(ŋ)gia	—	—	hiŋjō
stone		ko(m)bore[D, E]	—	—	kupuri
harbour		—	kunkudoi[I]	kukuruto[K]	—
low tide		—	siri[I]	tīna[J]	—
mud		—	meto[I]	matu[K]	—

[1] Buin *īto* has both nominal ('white smoke, white cloud') and verbal ('to be smoking') meanings. The variation in the final vowel is dialectal; *īta* is the nominal form in the Northern dialect (Laycock and Onishi 2003:25).

[2] Irregular lenition of *k as *w*.

Table 13: South Bougainville cognate sets: flora

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
betelnut	*mōti	mōsī[D]	moisi[I]	—	mosi[N], mōsi[P]
branch	*āgu[1]	āguʔ[D]	—	aku[J]	āŋ
coconut (tree)	*mou	mou, mō[D]	mou[I]	muo[2]	mō[M, P], mo[N]
fruit, seed	*tinaŋ	—	sinaŋ[H]	tiana[J, 3], siana[K]	sira[O, Q]
leaf	*pɔda	panda[D]	para	ota[J], hotahota[K]	hoto[P], patu[M. O]
mango	*baiti	was[D]	baisi[I]	paisi[K]	pah[P]
sugarcane	*tɔnɔŋ	tana[D]	tanaŋ[I]	—	toŋo[M, O], tono[O]
sweet potato	(*ane)[4]	ane[E]	ane[I]	āne[K], ane[L]	—
tobacco	*buru	—	burusi	pūtu, purutū[J, 5], purupūsu[K]	puruweku[N]
tree	*koi	koi[D]	koi	kui	kui
(tree) trunk	*mono	mono[D]	—	moruŋke[J], muruna[J, 6], munu[K]	munu[7]
arrowroot		—	—	kui polo[K]	kui poro[N]
bamboo		—	—	piti[J], pisi[K]	pih
banana (wild)		—	—	kourai[J]	kourai[M]
bark		—	—	kagu(a)[J]	kaŋ[N, P]
coconut (drinking)		—	—	kukutu	kukutu
sap		uriaʔ[D]	—	—	uroʔ[P]
yam		warah[D]	—	—	poro[P]
grass		musi[C]	—	mati[J]	—

[1] Compare cognate set for ‘hand’ in Table 9.

[2] *Muo* in Buin denotes both the coconut palm and the coconut fruit (Laycock and Onishi 2003:105).

[3] The form *tiana* is a Western Buin form, and although given as the equivalent for ‘fruit’, ‘seed’ and ‘nut’ appears to have the primary meaning ‘egg’ (Laycock and Onishi 2003).

[4] Sweet potato is a relatively recent introduction into New Guinea and the Pacific, and so this form most likely reflects the spread of a new lexical term rather than a form inherited from Proto South Bougainville.

[5] For the meaning ‘tobacco’, Laycock and Onishi (2003) list *pūtu* as the basic form, and *purutū* as a reduplicated variant.

[6] Buin *moruŋke* has the general meaning of ‘the central portion’, with ‘tree trunk’ listed as a secondary meaning. This form appears to be related go *moru* ‘middle, centre’. Laycock and Onishi (2003:106) also give the form *muruna*, which has the meaning of ‘trunk, stem of a tree’.

[7] Onishi (2002) glosses Motuna *munu* as ‘body, stalk, trunk’.

Table 14: South Bougainville cognate sets: fauna

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
bird	*bɔrege	ware(ŋ)ge	bareŋ	—	poriŋi
dog[1]	*masika	masika[C], mosika[D, E]	mosika[G, I], mosi-i[H]	(mairou, makirou)	mahkata
eel	*baramɔ	warama[D]	bara-ʔma[I]	palamo[K], maramo[J]	paramo[P]
fowl, chicken[1]	*kokore	kokorē[D]	kokore[I]	kukurei[J]	kukuraku
rat	*koto	kotōʔ[D], koso[C]	kutukai[H, I], kusikai[G]	kutukai[K]	kuʔtaki[M,N, P], kuhtaki[O]
louse		pasi	pansi[G, I], paŋsī[H]	—	—
flying fox		—	—	morokeŋ[J]	morokiŋ[M]
louse		—	—	oti[J], hosi[K]	hoh
fish		kare(ŋ)ge	—	—	koriŋi
mosquito		tāʔŋka[D]	—	—	tāŋka
turtle		—	mokono[I]	mukunu[J], mokono[K]	—

[1] Dogs and chickens were likely introduced by the Austronesians, and so these two cognate sets may post-date Proto South Bougainville.

Table 15: South Bougainville cognate sets: property terms

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
bad	*orara	orara[E]	orara	—	—
big	*pɔn(n)ɔ	panna[C, D]	pankaiŋ	oŋokopa[J]	honno[M, P]
black	*muŋ[i,o]	muni[C]	—	mūŋ[J]	mūŋo[M]
cold	*kamari	kamāri[A], kamali[C]	kamari[I], kamāri[G]	kamali[K], kamara[J]	kamaʔ[M], kamā[O]
dirty	*kumi	kumi(ʔ)[C, D]	kumi	kumi[K]	kuŋ
far	*iti-	isipo	isīpo	itigou[M], isigou[L]	ihko
hungry	*perɔ	pēransi[D]	pe-ʔra[I]	—	hīronai[M], hiro[P]
long, tall	*iti-	isikara[A, D], isikalo[C]	isikuŋ[G, I], isīkā[H]	iti-[J], isipa[K]	ihkita
old	*uri-	ulikala[C]	urika[H], urikung[G, I]	uni[J, 1]	—
sick	*tipɔ	sīpa	si-ʔpa[I]	hivo[K]	sīhowarei[M], siho[N, P]
thick	*mōtu	—	mōtu	mūtu	muhhā[O]
warm, hot	*tɔkɔtɔkɔ	takataka[D], takataha[C]	takatakā[G]	—	tokotoko[O]
white	*kākɔtɔ	kākata(ʔ)	kakara	kākata	kākoto
black		mutā[A]	mutāŋ[G, I]	—	—
blunt		motu[C]	motu[G, I]	—	—
full		mādo[D]	manto[I]	—	—
good		tabara[E]	tampara	—	—
new		nelakala	neraka[H], nerakuŋ[G, I]	—	—
red		ulugasi[C]	uruŋ[G, I]	—	—
rotten		lēra-[A]	dera	—	—
sharp		irido[D]	irinto[I], iriŋto[G, H]	—	—
dry		—	—	koporogasi[K]	koporagah[N, O], koporoŋarei[M]
new		—	—	riro[J]	dirokisa[N, P], nirokisa[O]
pain		masikēra	—	—	mōsika
straight		tono[D], tonaŋ[C]	—	—	tunupah
thin		nasikot[C]	—	—	nahkah[N]
sweet		—	tari-ʔ[I]	tori[K]	—

[1] The stem *uni* in Buin is a causative verb meaning ‘to make old, to make wise; instruct, teach’. This form also occurs as the first element in a number of lexemes with an apparent meaning of ‘old’. For example, *unigou* ‘old (place)’, with *-gou* is an adjectival suffix indicating ‘place’, *unimoko* ‘the past, olden times; long ago’, compare *moko* ‘descend’, and *unimokara* ‘a woman of former times, very old woman’ (Laycock and Onishi 2003:244, 19, 98).

Table 16: South Bougainville cognate sets: motion and stance

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
fall	*ru-	rura[E]	dua[I, G]	rū[J], nukale[K, L]	—
flow	*tū	—	tūkō[G]	tūra[J]	tūharei[N]
go	*be-	weʔ[D]	—	pe[J]	pinarei[M], pihe[P]
push	*tūme	tumele[C]	tūmē[H]	—	tūmiwarei[N], tumihe[O]
put	*ti-	sīgsi[E]	sinunpa[I]	ti[J], sisale[K,L]	sirarei[M,Q]
turn	*bero-	velo[C]	berekō[G], berebere[I]	pere[J]	—
come		pō[C], poʔnsi [D]	pokō[G]	—	—
pull		walasi[C]	barang[H, I]	—	—
stand		laŋ[C]	doŋkoŋkō[G], roŋoŋ[I]	—	—
come		—	—	u[J], huhale[K]	huharei[N], huhe[O, P, Q]
go in		—	—	tu[J], tukale[K]	tugarei[N]
go out		—	—	ti	sivale
lie on side		—	—	nau[J]	naukarei[M, N]
return		—	—	kāmuru[J]	kamurugarei[N]
run		—	—	kuro[J], kulogale[K]	kuroharei[N]
stand		—	—	ita[J], italale[K]	itarei[M, N], itihe[O]
walk		—	—	koŋ[J]	koŋkoŋ[N]
carry		ukansi	—	—	ukowarei[N]

Table 17: South Bougainville cognate sets: bodily functions and experiences

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
breathe, live	*roma-	lomasiaisi, lomaoto[C]	domantū[H], domāŋ[G]	numatu[J], lumatuale[K]	dumaruharei[N]
cough	*k(o)u-	—	kouʔ[K]	kugno[N]	—
die	*bō	wōʔnsi[D], vo[E]	bō(kō)[G, I]	pu(a)[J], puahale[K, L]	pū(harei)[M, P], puharei[N]
drink, eat	*nai	nai(nsi)[C, D]	nai[H, I]	rai, roi[J], noipale[K, L]	nēwarei[M, N, Q], neihe[O, P]
hear	*tarɔ-	tala(nsi)	tārakō[G], tara[I]	—	tarowarei[N, Q], taroihe[O]
sleep	*ati-	āsiʔ[D], asi[C, E]	āsikō[G], asi[H, I]	āti[J], asi[L]	ātarei[M], atarei[N], atihe[O], ātihe[P]
smell	*nū-	nū[C]	nū(kō)[G, H]	nūhale[K]	nuharei[N], nuhihe[O]
spit	*tutu-	—	tutupa-ʔ	turupai	turupagale, tūtūha
suck	*muti-	—	musimusi[I]	mutu[J], musukale[K]	musukarei[N], musukihe[O]
cry		vilokali[C]	biroŋ(kō)	—	—
vomit		kuli[C]	kuriŋ(kō)[G, I]	—	—
copulate		—	—	ruru[J]	rurukarei[M]
defacate		—	—	toro[J]	totorarei[M]
eat		—	—	pau[J]	paukarei[M, N], paururuhe[O]
heal		—	—	hula[L]	hura[Q]
stand		—	—	ita(lale)[J, K]	itarei[N]
vomit		—	—	kuŋ[J], kunsale[K]	kugnowararei[N], kunnawarihe[O]
awake		—	tante[I]	—	taniwarei

Table 18: South Bougainville cognate sets: basic actions and events

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
bite	*kɔki	kaki[C]	kakikō[G]	kaku[J], kakulale[K]	kokignarei[N]
buy	*bō-	wōri[D]	bo-ʔri[I]	pūŋ[J], punivale[K]	pūŋnarei[N], pū[L, 1]
cut	*tɔgi-	tagi[C, D]	tekakō[G], teʔka[I]	toki[J], tokisale[K]	tokisarei[M, N], tokisihe[O]
fear	*ɔro-	—	arorō[H]	ouru[J], ouluhoivale[K]	ōruharei[M], oruharei[N, O]
fly	*pure-	—	purereŋ(kō)[G, H]	ururu[J]	hurirarei[N]
give	*ɔb-	awena[C], auʔ[D], ausis[E]	abukō[G], abū[H]	o[J], ovale[K, L]	awarei[M], oihe[N] ōihe[O]
hit	*nip[o,i]-	—	nipo[I]	nipi[J], nipikale[K]	nihkarei[M, N], nihkihe[O]
kill	*tɔ-	tauʔnsi[D]	—	tā, to[J]	tawarei[M, N], toihe[O]
know, think	*ono-	—	onoukō[G], onouʔ[I]	ono[J], onohale[K]	onoharei[N, O], onohihe[O]
sew	*nɔti	nasi[C]	—	roti[J], nosihale[K]	noʔharei[N], nohihe[O]
shoot	*tō-	tōʔ[D]	tō(kō)	tua[J], suakale[K]	tōhēwarei[M], tohihe[O]
speak	*kɔrɔ-	karetāʔnsi[D]	karanau[I]	—	korokoro[M, P], korowarei[M, O, Q]
call		vokugsi[E]	boku[I]	—	—
find		lāʔnsi[D]	daʔ[I]	—	—
wash		lū	dū(kō)	—	—
ask		—	—	rakaro, nakaro[J], nakarovale[K]	nakarowarei[M]
call		—	—	pāro[J]	pāroŋarei[M]
forget		—	—	rore[J], loleuale[K]	roriharei
hold		—	—	tōro[J]	toʔkarei[M], tōkihe[O]
wash		—	—	ū[J], uvale[H]	ūharei[M, 2], uharei[N]

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
carry		ukansi[D], ukagsi[E]	—	—	ukowarei[P, Q]
fight		tauto[C, E]	—	—	tūtū[M], tutu(harei)[N, O, Q]
steal		tuna[E]	—	—	tūnawarei[N], tunaka[Q]
hunt		—	mānē[H]	—	mani
stab		—	tupu[I]	—	ruputarei[N]
swim		—	kuntu[I]	—	kunuʔharei[M]
hunt		—	mēkō[G]	mē[J]	—
split		—	bīŋsi[G]	pīsale[K]	—
wipe		—	piū[H]	hiukale[K]	—
ask for		oki[E]	—	uki[L]	—
throw		raro[E]	—	lolohale[L]	—

[1] Hall (n.d.) glosses Motuna *pū* as ‘to pay’.
[2] Onishi (2002) gives the gloss ‘bathe’ for Motuna.

Table 19: South Bougainville cognate sets: material culture

Gloss	Proto South Bougainville	Nagovisi	Nasioi	Buin	Motuna
canoe	*bakāti	—	bakasi[I]	pakāti[J]	hakāsi[M], hakasi, pakas[P,1]
house (general)	*papa	pawa	pava[I, 2]	opa[J], hopa[K]	howo
house (menstrual)	*bare	—	bare[I]	pale[K]	pari[N]
ornament (nose)	*napui	—	navui[I]	nāpui[K]	nawui[N]
warclub		—	—	pikupiku[J]	pikupiku[N]
axe (blade)		mareki[D]	—	—	mariki[P]
boat, ship		kaiboka[E]	—	—	kaipuka[N]
lime		maka[D]	—	—	moko[P]
netbag		wokata[D]	—	—	pokoto[P]
adze		—	teka[I]	—	tikaharei[N]
warclub		—	tavaka[I]	sapaka[K]	—

[1] While Nasioi *bakasi* and Buin *pakāti* are glossed as ‘outrigger canoe’, these Motuna forms appear to be generic terms for ‘canoe’.
[2] The Nasioi form for ‘house’ in Grace and Davika (1955) is not entirely clear.

6 *The languages of Vanikoro: three lexicons and one grammar*

ALEXANDRE FRANÇOIS

> Cette île, toute petite qu'elle est, présente le singulier phénomène de plusieurs idiômes differens. (Gaimard 1833:338)

1 The paradox of Vanikoro languages[1]

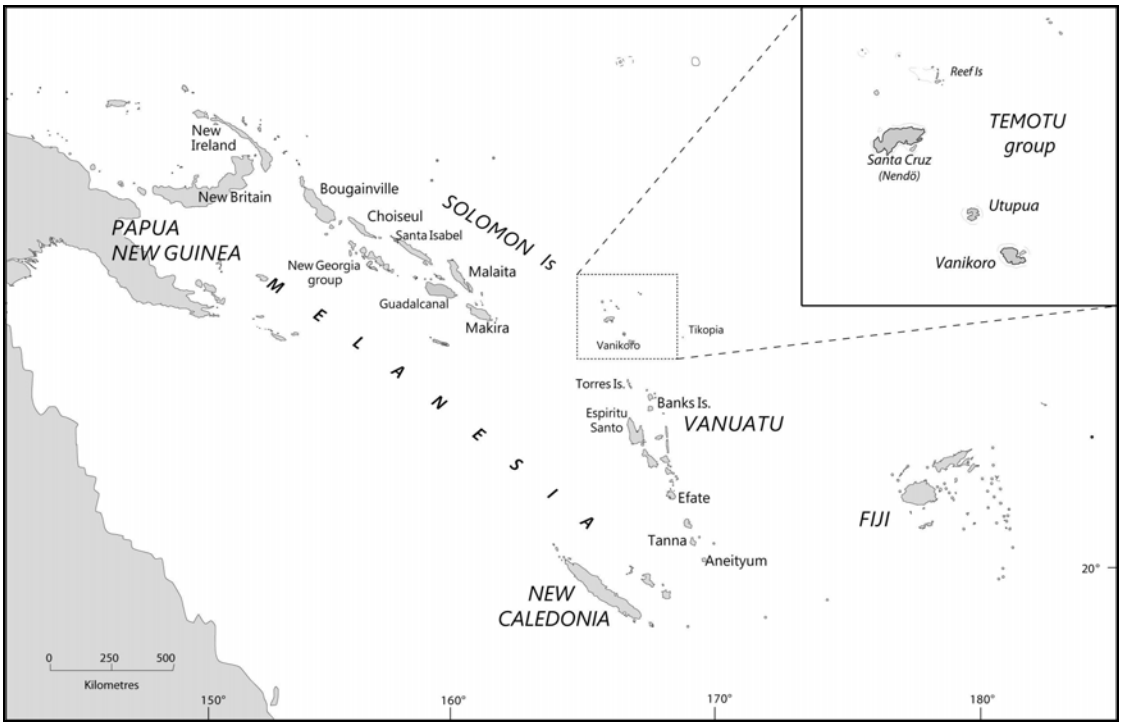
1.1 The languages of Vanikoro

With its 193 sq. km, Vanikoro is the second largest island in the small archipelago formerly known as the Santa Cruz Islands, and now often referred to as 'Temotu', after the official name of the easternmost province of the Solomon Islands (Map 1).

The province of Temotu is home to a variety of languages (Tryon 1994): three Polynesian, and nine non-Polynesian. The latter include three languages on Vanikoro, and three on Utupua — a total of six Oceanic languages which have long been understood to form a branch of their own (Tryon and Hackman 1983). The three remaining languages, known as the 'Reefs-Santa Cruz', were long deemed to be Papuan (Wurm 1976), but have recently been shown to be Austronesian (Ross and Næss 2007). More specifically, Ross and Næss have proposed to group all the non-Polynesian languages of the region into a single first-order subgroup of Oceanic, labelled 'Temotu'. The latter would then split into two branches: Reefs-Santa Cruz (RSC) on the one hand, and Utupua-Vanikoro (UV)[2] on the other hand.

[1] My gratitude goes to Malcolm for having fostered my linguistic research on Melanesian languages, both through his writings and through our discussions, ever since we first met in 1997. His interest in my data has been a strong incentive for me to spend the years 2009 and 2010 at The Australian National University. My initial work on Vanikoro was facilitated by the Institut de Recherche pour le Développement of Nouméa and Association Salomon; by the French Centre National de la Recherche Scientifique; by Piet Lincoln; by Association 'Banie' and the traditional chiefs of the island; and by my Vanikoro helpers and friends — especially Stanley Repuamu, Ezekiel Prians, John Nabu. I am grateful to Bethwyn Evans and Andrew Pawley for their comments on earlier versions of this paper.

[2] Except for its new location in the POc tree, this UV branch coincides with the subgroup identified earlier as 'Eastern Outer Islands' (Tryon 1994, 1995).



Map 1: Location of Vanikoro in the Pacific

If Ross and Næss' hypotheses are correct, then Temotu constitutes a new branch of the Oceanic family tree whose history needs to be written. The present paper hopes to play its part in this endeavour, by presenting primary data and some discussion on the three indigenous languages of Vanikoro: Teanu (TEA), Lovono (LVN) and Tanema (TNM).

Published information on Vanikoro languages is still limited, but certainly not absent. In 1788, the island of Vanikoro was reached by the two frigates of the French navigator Jean-François de La Pérouse, and saw his fatal shipwreck — an event which was only understood a few decades later by the Irish navigator Peter Dillon (Dillon 1830). Another French officer, Dumont d'Urville, immediately organised an expedition in Dillon's wake. Among the abundant documentation produced by this second voyage (Dumont d'Urville 1830–1834), the French naturalist Gaimard (1833, 1834) compiled wordlists in the three languages of Vanikoro, a precious document on the linguistic situation of that time.[3] Much more recently, other word lists have been compiled by Tryon and Hackman (1983), based on an extended version of the Swadesh basic vocabulary list. Darrell Tryon also wrote short grammatical accounts of Lovono (Tryon 1994:630–634), and of Teanu (Tryon 2002). Additionally, a short collection of Teanu texts was published in Tua and Lincoln (1979).[4]

Almost two centuries after Dumont d'Urville's expedition, the French Ministère de la Marine, together with Association Salomon and Institut de Recherches pour le Développement, organised another expedition called *Vanikoro 2005*, to find out about the fate of La Pérouse's ships and sailors. I was given the opportunity to play my part there as

3 I am much indebted to Piet Lincoln for allowing me easy access to Gaimard's documents.

4 The two languages Teanu and Lovono have been given varying names over time. Teanu was called *Tanéanou* by Gaimard, and *Buma* by Tryon. Lovono was called *Vanikoro* by Gaimard, *Vanikolo* by Ivens (1918), *Vano* by Tryon. See §1.2 below for a discussion of my naming proposals.

a linguist, documenting place-names and oral traditions, with a special interest in the islanders' stories that still remember so vividly the 1788 wreckage (François 2008a). On this occasion, I was also able to follow in Gaimard's footsteps, and record what I could of the three languages. One thing I realised was the urgency of this task, with both Lovono and Tanema remembered by only a handful of speakers.

1.2 A note on the history and geography of Vanikoro

The population of Vanikoro can be described at two different levels of observation.

A contemporary look would probably suggest just a binary divide between two communities, one Melanesian and one Polynesian. The latter is a group of about 300 settlers originating from Tikopia, a small island located about 200 km eastwards. Although they have been colonising the southern shores of Vanikoro for more than three centuries (see Dillon 1830), they tend to interact very little with the native population — except for the occasional land dispute. As their social network is still anchored in their Polynesian homeland, they remain predominantly monolingual in Tikopian, the Polynesian Outlier language spoken on Tikopia (Firth 1985). Apart from a few loanwords here and there, this recent colonisation does not show any major linguistic consequence, and will not be discussed further.

With about 600 individuals descending from the earlier inhabitants of Vanikoro, the Melanesians live today in six coastal villages scattered around the island: Puma, Temuo, Paiu, Lale, Lovono, Lovoko. People from these villages communicate through coastal canoeing, and now form a homogeneous society. This modern unity reportedly results from the action of the Anglican missionaries who christianised the region at the end of the 19th century. However, the first historical documents on Vanikoro (Dillon 1830; Dumont d'Urville 1830–1834), as well as the oral tradition of the islanders themselves, tell a different story: that of an island which used to be sharply divided into three distinct 'tribes' or chiefdoms. Each of these tribes was attached to a specific territory which they defended fiercely from one another, and which is still clearly delimited in people's memories, see Map 2.

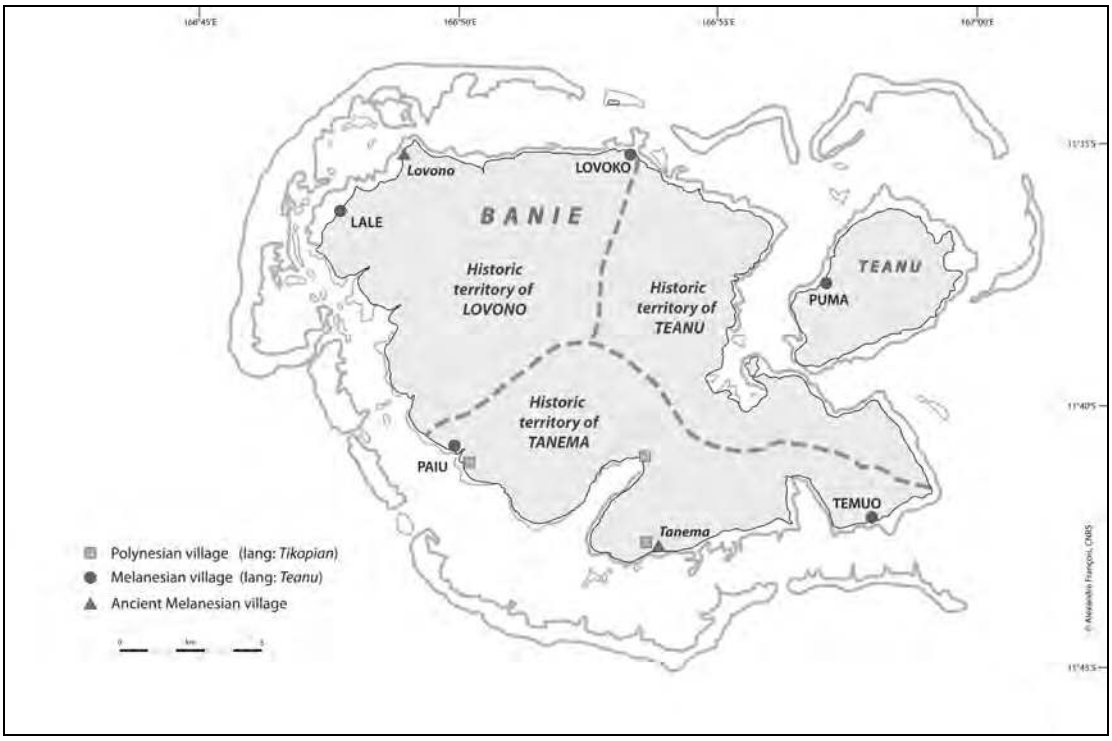
Each tribe's name recalls a significant place of its own area:

- ***Teanu*** from the name of the northeast island of the Vanikoro group, where the village of *Puma* is also located[5]
- ***Lovono*** from the name of a village, also formerly known as *Vono* or *Vano*,[6] northwest of the main island Banie
- ***Tanema*** an ancient village, also known as *Tetawo*, on the southern coast (in what is now *de facto* Polynesian territory)

---

[5] The village name *Puma*, sometimes with the incorrect spelling *Buma*, has been used as another name for the language of Teanu (see footnote 4).

[6] The Lovono name of this village of Vano/Lovono was apparently *Vana* or *Alavana*. In principle, one might want to use this autonym as a reference name for this language, yet this would be slightly artificial: the only term which is used today, even by the last speakers themselves, is the form *Lovono* in Teanu, the only surviving language.



Map 2: The populations of Vanikoro

Each of these three tribes had its own language, which survived up until the 20th century. Gaimard's (1833, 1834) wordlists cite 'Tanéma' and 'Vanikoro' (Lovono) on an equal footing with 'Tanéanou' (Teanu). Ivens (1918:155), in his translations of the prayer 'Our Father', illustrates Vanikoro with a language that can be identified as Lovono — which suggests it was still in a healthy state at the end of the 19th century. But since the pacification of the island, intermarriage amongst the three earlier communities has increased, and they have merged into a single society. Probably due to its earlier demographic lead, Teanu was adopted as the whole island's daily language, very much at the expense of the two other vernaculars. In 2005, Lovono and Tanema were still remembered by only a handful of elder speakers — five for Lovono, four for Tanema. Because these last speakers are now dispersed across the various Teanu-speaking villages of Vanikoro, they do not form anything that would resemble a speech community. Clearly, the shift to Teanu is now complete: Lovono and Tanema are now two moribund languages, with only a few years left to live.

1.3 Three closely related languages

1.3.3 Three aberrant Oceanic languages

For the linguist accustomed to other Oceanic-speaking areas, the three languages of Vanikoro can be disconcerting. This section will touch briefly upon the question of their genetic affiliation, before we examine the issue of their mutual relationship.

One conspicuous characteristic of Teanu, Lovono and Tanema is the degree to which they differ, both lexically and morphologically, from most other known Oceanic languages (François 2006), whether from the Solomons or from nearby Vanuatu. For example, these

three languages show relatively little retention of the lexicon reconstructed for Proto Oceanic. Many etyma, which tend to be otherwise widely preserved among Oceanic languages (Pawley 2007), have disappeared altogether from the lexicon of Vanikoro languages; e.g. *tama- 'father', *mate 'die', *kani 'eat', *kayu 'tree', *ikan 'fish', *sapa 'what', *jalan 'road', *susu 'milk', *pano 'go', *sake 'up', *sipo 'down'. The morphology also shows a great deal of erosion. Thus the languages of Vanikoro show no trace whatsoever of the POc possessive affixes (*-gu, *-mu, *-ña…; *-qi…), of object pronouns, of the article *na, of common verbal affixes like the transitive suffix *-i, the applicative *-aki[n], the causative *pa[ka]-, and so on.

In sum, Vanikoro languages can be considered 'aberrant' (see Grace 1990; Pawley 2006) in comparison with most other Oceanic languages. This does not mean, however, that their Oceanic ancestry cannot be detected. Indeed, despite the high level of lexical replacement which evidently took place in their history, it is still possible to uncover some words whose similarity with Oceanic reconstructions is beyond doubt, and which could hardly be attributed to chance or borrowing — if only because their phonological correspondence patterns seem to follow some degree of regularity. Table 1 provides a small sample list of such words.[7]

Table 1: Some obvious Oceanic reflexes across the three languages

English	Teanu	Lovono	Tanema	POc etymon
'bird'	*menuko*	*menuka*	*manuke*	*manuk
'eye'	*mata*	*mala*	*ka\mae*	*mata
'soul, spirit'	*ata*	*ala*	*ae*	*qata
'ears'	*tañа*	*mabe\leŋe*	*añe*	*taliŋa
'house'	*moe ~ mwoe*	*moe*	*nalama*	*ʀumaq
'cold'	*medigo*	*meniŋa*	*mediŋa*	*ma$^{(n)}$ri$^{(n)}$riŋ
'long, tall'	*biouro*	*beure*	*va\beura*	*barapu
'Canarium nut'	*vo\ŋoro*	*ve\ŋere*	*vi\ŋara*	*[ka]ŋaʀi
'lie down'	*wene*	*enu*	*eno*	*qenop

1.3.2 Proto Vanikoro, the common ancestor

Not only can Teanu, Lovono and Tanema be individually linked to Proto Oceanic, but they can also be shown to form a set of three closely related languages. It is possible to suggest instances of exclusively shared innovations, thereby pointing to the likely existence of a common ancestor Proto Vanikoro. The following paragraphs therefore answer the question raised by Ross and Næss (2007:473), according to whom 'no innovations define Vanikoro'.

---

[7] The three Vanikoro languages have the same phoneme inventory, with five short vowels (i, e, a, o, u) and 19 consonants. The spelling conventions here adopted include *v*=/β/; *ñ*=/ɲ/; *j*=/ ^{n}ʒ/. Labiovelar consonants use digraphs with a *w*, and all prenasalised voiced stops are spelled without their nasal element, e.g. /^{m}b^{w}/ is *bw*, /^{ŋ}g/ is *g*, etc.

In the phonological domain, one can cite the addition of a paragogic vowel, probably a schwa, after most word-final consonants, resulting in the retention of that consonant followed by a non-etymological vowel:

(1) POc *manuk 'bird' > *manuk**ə** > TEA *menuk**o*** ~ LVN *menuk**a*** ~ TNM *manuk**e***;
POc *ma^{n}ri^{n}riŋ 'cold' > *ma^{n}riŋ**ə** > TEA *medig**o*** ~ LVN *meniŋ**e*** ~ TNM *mediŋ**a***;
POc *waiʀ 'water' > *waiʀ**ə** > TEA *er**o*** ~ LVN *wir**e*** ~ TNM *n\ir**a***.

Another example of a sound change which is only found in Vanikoro languages is the occasional velarisation of *t to /k/ before a high back vowel /u/:

(2) POc *patu 'stone' > *vatu > *vaku > *vakə > TEA *vok**o*** ~ LVN/TNM *vak**a***;
POc *kutu 'louse' > *utu > *uku > *ukə > TEA *uk**o*** ~ LVN/TNM *-uk**a***.

In the morphological domain, we will see (§3.1.2) that the three languages share the same structural collapse between certain non-singular personal prefixes.

Finally, many lexical items seem to be shared exclusively by these three languages. Table 2 proposes a set of possible lexical innovations in Proto Vanikoro — with tentative reconstructions of the most likely etymological forms.[8]

Table 2: Some putative lexical innovations in Proto Vanikoro

English	Teanu	Lovono	Tanema	Proto Vanikoro
'moon'	*metele*	*mele*	*maloula*	*mataul(ə)
'fish'	*namuko*	*namweka*	*namaka*	*nam$^{(w)}$əkə
'wood, tree'	*vilo*	*kuile*	*veila*	*v$^{(w)}$eilə
'taro'	*je\bute*	*bule*	*bue*	*bute
'child'	*menu*	*melika*	*anuka*	*menuk(ə)
'woman'	*emele*	*neme*	*me*	*nemel(ə)
'name'	*eŋa*	*neŋe*	*niŋe*	*neŋə
'who'	*ŋele*	*eŋe*	*ŋela*	*eŋel(ə)
'inhabited land'	*kulumoe*	*kulamoe*	*kulama*	*kuləmaʔ(ə)
'rubbish'	*ajekele*	*togale*	*soge*	*jogel(ə)
'tell s.o.'	*viñi*	*puñi*	*puña*	*puñə
'heavy'	*motoro*	*melure*	*mwaura*	*matur(ə)
'dirty'	*sukiro*	*sukure*	*vatukura*	*sukir(ə)
'stand'	*vio*	*pia*	*veo*	*piə
'impede, protect'	*botoŋo*	*beloŋa*	*baoŋa*	*batoŋə
'do again'	*tabo*	*lebu*	*abo*	*tabo

[8] Due to the lack of solid regularity, in particular, in the domain of vowels, it is difficult to securely reconstruct any protoform based on the synchronic data. This is an important difference between the languages of Vanikoro and other parts of Oceania, where protoforms can be reconstructed based on modern reflexes. See Ross (1988) for the languages of western Melanesia, Lynch (2001) for southern Vanuatu, François (2005) for northern Vanuatu.

In order to ascertain these reconstructions and expand the list of Proto Vanikoro innovations, more needs to be known of the languages on the neighbouring island Utupua. But the evidence given here should be sufficient to establish that the three Vanikoro languages form a subgroup of their own, pointing to a single common ancestor.

1.4 Divergence and convergence among Vanikoro languages

In sum, Teanu, Lovono and Tanema are three 'aberrant' Oceanic languages which are genetically closely related to each other, as descendants of a single ancestor. Now as we compare the three modern languages, we may be surprised by a paradox:

- Compared to the situation in most other Oceanic languages, the forms of words in the three languages of Vanikoro tend to differ from each other in ways which can be regarded as quite extreme given their geographic closeness.
- Despite their heterogeneity with regard to word forms, they show perfect isomorphism of their structures.

This paradox will be the main focus of this article.

Throughout this paper, I will refer to two essential components of language, which crosscut the traditional division between lexicon and grammar. On the one hand, I will mention STRUCTURES, referring to the various concepts and semantic categories with which a language divides up semantic space — whether this refers to lexical or to grammatical meaning. On the other hand, each language embodies these categories and concepts into linguistic FORMS, endowed with a specific phonological content.[9] Two languages can be said to be isomorphic in a particular domain of their system, if they share the same structures or semantic categories, whether or not the forms they use are cognate with each other. For example, the two constructions *I have caught cold* and *J'ai attrapé froid* are perfectly isomorphic, because they express the same event by resorting to exactly identical metaphors and categories.

My observation is that the three Vanikoro languages exhibit a high degree of divergence in their forms, yet still show an extreme isomorphism of their structures. This configuration is illustrated in example (3). As far as the grammar is concerned, the three languages possess parallel structures and word order, to the point that they can all be analysed with a single line of word-to-word glosses. Yet on the other hand, one can equally note the dissimilarity between the actual forms of their words.[10]

(3)

	TEA	*A-ko*	*u-ka*	*u-katau*	*ene ?*
	LVN	*Nu-pu*	*ku-ma*	*ku-ki*	*ŋane ?*
	TNM	*Go-po*	*go-loma*	*go-ie*	*nana ?*
		2SG:R-say	2SG:IR-come	2SG:IR-follow	1SG

'Do you want to come with me?'

---

9 This contrast STRUCTURES versus FORMS corresponds to what Hjelmslev (1961:52) described as respectively *content form* versus *expression form*; and to what Grace (1981:24) would call *content form* versus *lexification*.

10 Abbreviations in glosses include: R - realis prefix; IR - irrealis prefix; GEN - General possessive classifier; INDEP - independent pronoun; FOOD - possessive classifier for food possession; HUM - article for human referents.

This observation, whereby languages can be at once homogeneous in structure while dissimilar in forms, has been widely made in the linguistic literature already (see Gumperz 1971; Enfield 2001). However, most of the time, those facts of structural parallelism result historically from contact between genetically diverse languages. To focus on works on the Melanesian area, Thurston (1989, 1994) thus describes the structural similarities between languages of northwestern New Britain, some Papuan and some Oceanic, and Ross (1996, 2001) discusses the influence of Waskia (Papuan) upon Takia (Oceanic), in Papua New Guinea's Madang Province. To this list, one could add discussions of structural parallelism between the English-based pidgins of the Pacific and their Melanesian substrates (Camden 1979; Keesing 1988, 1991; Siegel 2008). In all these cases, whether one compares Oceanic with Papuan or with European languages, the diversity of forms is a given; and what is observed is first and foremost a matter of STRUCTURAL CONVERGENCE — or 'metatypy', to use the term coined by Malcolm Ross (1996, 2001).

But the case of Teanu, Lovono and Tanema raises different issues, because they belong to the same genetic subgroup. For such closely related languages to share the same syntactic structures may partly reflect the mere legacy of their common ancestor, and partly be explained by later contact-induced convergence. The intriguing part here is rather the dissimilarity of forms: it needs to be considered not as a simple given — as was the case for genetically diverse languages — but as the problematic result of historical DIVERGENCE from a common ancestor. This configuration therefore requires specific explanations, beyond the now well-known cases of metatypy.[11]

Section 2 will discuss the degree of similarity and dissimilarity existing between the lexical forms of Vanikoro languages. Section 3, in turn, will demonstrate their strong structural isomorphism. Section 4 will finally propose a functional hypothesis to account for this linguistic paradox of Vanikoro languages.

2 Similarity and divergence of lexical forms

The linguistic relations between the three Melanesian languages of Vanikoro can thus be summarised in a simple formula: DISSIMILAR FORMS, SIMILAR STRUCTURES. The present §2 will discuss the first of these two dimensions, by assessing the degree of differentiation between the lexicons of Teanu, Lovono, and Tanema.

Even though Teanu, Lovono and Tanema are close genetic relatives (§1.2), the impression that prevails is that of a rather strong dissimilarity of their lexicons. Indeed, contrary to the impression given by Table 1 above, the three languages of Vanikoro are not mutually intelligible. Their dissimilarities are of varying nature, going from more or less regular phonological differences between cognate words, to forms that are simply non-cognate. Overall, while such formal differences are frequent in the Melanesian area, their degree is here rather impressive for languages which are spoken on the same island, and appear to have historically differentiated on this island.

---

[11] A similar blend of formal dissimilarity and structural parallelism can be found in the languages of north Vanuatu (François 2007, forthcoming, in prep.). However, the phenomenon appears to be even more conspicuous in the case of Vanikoro.

2.1 Differences due to phonological change

In some cases, forms which are superficially very dissimilar can in fact be explained by regular processes of sound change. While many correspondences between the three languages of Vanikoro are straightforward and obvious, some are more drastic and can result in little resemblance between the actual forms.[12]

For example, the verbs for 'sit, stay' (TEA *te* ~ LVN *lu* ~ TNM *o*) are dissimilar enough to suggest they might be non-cognate. However, one can establish a regular correspondance pattern TEA /t/ ~ LVN /l/ ~ TNM Ø (see sample in Table 3) pointing to a proto-consonant *t (Tryon and Hackman 1983:71).

Table 3: Some lenis reflexes of POc *t

English	Teanu	Lovono	Tanema	POc etymon
'sit, stay'	***t****e*	***l****u*	*o*	*toka
'three'	*te-**t**e*	*te-**l**u*	*a-o*	*tolu
'sugarcane'	***t****o*	***l****epie*	*ova*	*topu
'soul, spirit'	*a**t**a*	*a**l**a*	*ae*	*qata
'ghost, spirit'	***t****adoe*	***l****eñoe*	*aoe*	*qata- [?]
'do again'	***t****abo*	***l****ebu*	*abo*	
'carry on shoulders'	***t****abe*	***l****ebe*	*ebe*	
'impede, protect'	*bo**t**oŋo*	*be**l**oŋa*	*baoŋa*	
'unripe, new'	*mo**t**oe*	*me**l**oe*	*ma**j**a*	*mataq

The three forms for 'sit, stay' therefore suggest an etymon *tV. While correspondences are much less obvious regarding vowels, a proto-form with /o/ is a likely origin for a pattern TEA /e/ ~ LVN /u/ ~ TNM /o/; see also the vowels of 'lie down' (< POc *qenop) in Table 1, and of 'three' in Table 3. The perfectly parallel reflexes for 'three' (< *to < POc *tolu) and for 'sit, stay' confirm a reconstruction *to — certainly the first syllable of POc *toka 'stay'. Interestingly, Gaimard (1833, 1834) consistently writes these Tanema words with an *r*, which shows the correspondence pattern was TEA /t/ ~ LVN /l/ ~ TNM /r/ two centuries ago. Thus, he writes *rarou* for 'three' (modern *a-o*), and *guidiro* for what he glosses '*Asseyez-vous*'; the latter in fact representing **giti-ro* (modern *giti-o*) 'we$_{[INCL]}$ are sitting'.

In sum, the regularity of correspondences, when they can be established, makes it possible, quite classically, to detect the cognacy of some forms which would have otherwise seemed unrelated.

Sometimes, the ultimate POc source of a given series is unclear, yet at least one can tentatively draw connections between modern forms, based on synchronic regular sound correspondences. For example, the verb for 'die' is *bu* in Teanu and Tanema, and *me* in Lovono, two forms with little in common. However, a regular sound pattern seems once again to emerge from the data: TEA /b/ ~ LVN /m/ ~ TNM /b/ — with a small amount of variation involving voicing or rounding of the consonant (Table 4).

[12] Some of the regular correspondences are given in Ross and Næss (2007).

Table 4: A possible regular correspondence pattern

English	Teanu	Lovono	Tanema	POc etymon
'die, dead'	***b**u*	***m**e*	***b**u*	(*mate) ?
'sharpen, trim'	***b**o*	***m**e*	***b**o*	
'1exc:dual pronoun'	*ke**b**a*	*ge**m**a*	*ga**b**e*	*kama[m]i
'k.o. basket'	*iunu**b**o*	*nunu**m**ie*	*nu**b**a*	
'tobacco'	*na**b**ene*	*naka**m**ene*	*na**bw**ane*	
'blood'	*a**b**o*	*a**mw**ale*	*a**b**a*	
'Areca catechu'	***b**uioe*	*na**mw**e*	***b**uia*	*buaq
'Reflexive-Reciprocal'	*ñe**p**e*	*ñe**m**e*	***b**e*	
'fishing net'	***p**ele*	***m**enele*	***b**enala*	*kup^{w}ena ?

This suggests the three forms for 'die' may be cognate, despite their present dissimilarity. In this case, because no known POc reconstruction (including *mate) provides any satisfying etymon, the cognacy judgment rests on purely synchronic data.

Table 5: Some potential, but dubious, cognate sets

English	Teanu	Lovono	Tanema	POc etymon
'man, person'	*mwaliko*	*lamuka*	*anuka*	*m^{w}aqane ?
'women'	*viñevi*	*veñime*	*?*	*pine ?
'canoe, ship'	*kuo*	*nawe*	*goia*	*waga(ŋ) ?
'sleep'	*mokoiu*	*mepeu*	*matou*	*maturuʀ ?
'red'	*moloe*	*wamoene*	*manobeila*	*meʀaq ?
'randomly; in vain'	*moli*	*moli*	*mano*	
'help; with'	*samame*	*emeio*	*avaio*	
'where?'	*vele*	*mane*	*vane*	
'I, 1sg pronoun'	*ene*	*ŋane*	*nana*	
'sink'	*metelu*	*mwelesu*	*madilo*	
'perhaps'	*bwara*	*bweti*	*buru*	
'go down'	*abu*	*pwo*	*kabu*	
'good'	*wako*	*vakane*	*apika*	
'seize, hold'	*labu*	*lo*	*nou*	
'rejoice'	*pei*	*pwadi*	*pae*	
'stone oven'	*awene*	*epene*	*pavene*	

Occasionally, the modern forms display little more than a vague 'family resemblance'. That is, the modern lexical forms are possibly cognate, yet their phonemes enter no regular correspondence pattern, in such a way that one could only explain their cognacy by resorting to *ad hoc* etymological hypotheses. Table 5 provides a sample of such potential, but irregular and dubious, cognate sets, which would require closer scrutiny in the future.

2.2 Lexical replacement

Finally, it also often happens that the three languages have forms that are distinct, and almost certainly not cognate for the same meaning. A sample of such cases is given in Table 6.

Table 6: Some non-cognate sets resulting from lexical innovation

English	Teanu	Lovono	Tanema
'thing'	*ŋatene*	*vesemele*	*vamora*
'know'	*ovei*	*lonei*	*wo*
'come'	*ka*	*mage*	*loma*
'lie, deceive'	*tomoli*	*ñaine*	*role*
'see'	*romo*	*eti*	*runi*
'quickly'	*kiane*	*segei*	*gamoi*
'bad'	*tamwaliko*	*visale*	*vae*
'big (PLUR)'	*wopine*	*evala*	*bwau*
'broken'	*mamakoe*	*seli*	*vave*
'remote'	*somu*	*akaole*	*mosomu*
'down, below'	*puo*	*lenu*	*ese*
'what?'	*(ŋan)ae*	*ese*	*sive*
'do what, do how?'	*(mi)kae*	*ñese*	*jive*
'be why?'	*ve*	*wo*	*ja*
'another, an'	*iote*	*leka*	*keo*
'one; same'	*iune*	*tilioko*	*omwano*
'play; wander'	*moloe*	*telu*	*lumota*
'neck; mind'	*awa*	*warene*	*vasare*
'bush'	*ŋogoro*	*atere*	*arara*
'cabbage'	*tebo*	*lamware*	*some*
'rat'	*uvilo*	*katone*	*ivala*
'be ripe'	*ako*	*wi*	*kou*
'guts'	*bea*	*lale*	*lebwe*
'year'	*ebieve*	*verue*	*rove*
'chicken'	*kulevelu*	*kio*	*tokila*
'leaf'	*uie*	*nugia*	*lele*

In their survey of Solomon Island languages, Tryon and Hackman (1983:481) give the following cognate percentages for the three languages of Vanikoro, based on a modified version of Swadesh's basic vocabulary list (200 words):

- Teanu–Lovono 57.3 %
- Teanu–Tanema 51.1 %
- Lovono–Tanema 54.8 %

These figures point to a relatively high level of lexical dissimilarity. They surpass similar counts made in most other parts of island Melanesia. For example, the Torres and Banks Is

of north Vanuatu are another area where lexical replacement has been intense (François, in prep.). However, in order to find figures as low as those on Vanikoro, one has to pick languages which are geographically spread apart; e.g. 44.5 % between Hiw (Torres) and Lakon (Gaua, south Banks). Should one consider languages spoken on a single island, the widest gap one can find there is between Lakon and Dorig, with 61.5 % shared vocabulary (Tryon 1976:95).

Even more instructive is the comparison with other language families in the world. To take just one example, rates of shared vocabulary amongst Germanic languages do not go below 53.6% (Dyen, Kruskal and Black 1992). In other words, the three languages of Vanikoro have managed to achieve, within the limited space of a single island, more lexical diversity than the whole Germanic subgroup of Indo-European. Such an intense differentiation deserves to be acknowledged, and discussed (see §4).

The strong formal dissimilarity that prevails between the three languages of Vanikoro naturally results in sentences where the phonological form of words, whether lexical or grammatical, can show a high degree of dissimilarity:

(4)

TEA	*Pi-te*	*ne*	*sekele*	*iupa,*	*pi-wowo*	*uo.*
LVN	*Nupe-lu*	*ne*	*amenoŋa*	*iemitore,*	*nupe-ŋoa*	*upie.*
TNM	*Tei-o*	*ini*	*vasaŋola*	*akegamuto,*	*ti-oa*	*uva.*
	1EX.PL:R-stay	in	garden	our	1EX.PL:R-plant	yam

‘We were in our garden, we’ve been planting yams.’

(5)

TEA	*Somu*	*tamwase*	*tae,*	*vitoko*	*takoie*	*ne.*
LVN	*Akaole*	*visalewabeu*	*taie,*	*vateoko*	*ŋate*	*ida.*
TNM	*Mosomu*	*vaepamabo*	*eia,*	*vatako*	*eto*	*kana.*
	remote	very	NEG	close	uphill	here

‘It is not very far; it’s close to here, up this way.’

The impression of formal heterogeneity may be due partly to sound change affecting words that are in fact cognate (e.g. ‘yam’, ‘stay’, ‘plant’); and partly to lexical replacement (e.g. forms for ‘very’, ‘remote’, ‘uphill’) or morphological change (e.g. forms of personal pronouns).

What is perhaps more puzzling is the contrast between, on the one hand, this formal dissimilarity, and on the other hand, the perfect parallelism existing between these languages’ structural properties. This is the topic of the next section.

3 Structural isomorphism

The push towards linguistic differentiation has affected the phonological forms of words, yet evidently had little impact on grammatical and semantic structures. Even as their lexicons were diverging from each other, the three languages of Vanikoro have maintained a strong STRUCTURAL ISOMORPHISM — whether in syntax, phraseology or organisation of meaning. This is reflected, for instance, by the ability of translating word-for-word any sentence from one language to another, following the same word order and the same morphological and lexical categories — as in examples (3) to (5).

The present section will illustrate this strong structural parallelism using a few conspicuous examples from various aspects of the grammar, and will end with a tentative account of the linguistic history of Vanikoro.

3.1 Morphosyntax

3.1.1 Syntax of the clause

Teanu, Lovono and Tanema are parallel in all aspects of their syntax. This includes all properties related to word order, whether the basic SVO clause order, or other properties, use of prepositions, post-nominal placement of adjectives and possessors, clause-final position of the negation as in (5).

The three languages display the same organisation in parts of speech, including a sharp divide between nouns and verbs, and a distinctive category of adjectives. While all verbs are obligatorily prefixed for subject and mood (§3.1.2), adjectives are unprefixed. Languages even agree on whether they treat a specific notion as an adjective or as a verb, as evidenced by the parallel presence vs absence of the subject prefix in the two predicates of (6).

(6)

TEA	*Udo*	*ponu,*	***boro***	*we*	*i-**ako***	*?*
LVN	*Puŋa*	*pae,*	***bware***	*we*	*i-**wi***	*?*
TNM	*Uda*	*pade,*	***betika***	*we*	*i-**kou***	*?*
	banana	that	black/unripe	or	3SG:R-be.ripe	

'Those bananas, are they green$_{[ADJ]}$ or ripe$_{[VERB]}$?'

Likewise, all syntactic properties of the clause are reflected alike across the three languages. They all lack noun articles, case markers, verb transitivisers or applicatives, and causative affixes. They make use of a reflexive marker (Table 4), which is also a reciprocal and an emphatic marker. They all resort frequently to core-layer verb serialisation, with exactly parallel phrasing; see examples (3), (7) and (10). They possess exactly parallel Tense-Aspect-Mood categories, and so on.

3.1.2 Pronouns and TAM marking

The paradigms of personal pronouns are also organised in parallel ways. Like most Oceanic languages, those of Vanikoro distinguish between exclusive and inclusive 'we'; and they show three numbers: singular, dual, plural.

Table 7 shows the eleven independent pronouns for the three languages. These show a reasonable degree of similarity.

Table 7: Personal independent pronouns

	Teanu	Lovono	Tanema
1sg	ene	ŋane	nana
2sg	eo	ago	go
3sg	ini	ŋani	nini
1in:du	kia	gita	gie
1ex:du	keba	gema	gabe
2du	kela	gamila	gamile
3du	da	dea	delalu
1in:pl	kiapa	gitu	geto
1ex:pl	kupa	gamitu	gamuto
2pl	kaipa	gaipa	gamito
3pl	dapa	detu	dato

Besides these free pronouns, these languages also possess[13] a double set of mood-marked subject prefixes for verbs (one for realis, one for irrealis); see Table 8.

Table 8: Verbal prefixes for subjects

	Teanu		Lovono		Tanema	
	Realis	Irrealis	Realis	Irrealis	Realis	Irrealis
1sg	ni-	ne-	ni-	ka-	ne/i-	na-
2sg	a-	u-	nu-	ku-	go/i-	go-
3sg	i-	i-	i-	ki-	i-	i-
1in:du	la(i)-	la(i)-	la(i)-	sa-	de-	ja-
1ex:du	ba(i)-	ba(i)-	(nu)ba-	ba(i)-	ba(i)-	ba(i)-
2du	ba(i)-	ba(i)-	(nu)ba-	ba(i)-	ba(i)-	ba(i)-
3du	la(i)-	la(i)-	la(i)-	sa-	de-	ja-
1in:pl	li-	le-	le(pe)-	kape-	le/i-, giti-	la-
1ex:pl	pi-	pe-	nupe-	pe-	te/i-	tu-
2pl	pi-	pe-	nupe-	pe-	te/i-	tu-
3pl	li-	le-	le(pe)-	se(pe)-	le/i-	la-

Beyond their general family resemblance, the subject prefixes shown in Table 8 show a certain amount of formal variety, see especially the 2sg, or the plural forms. Yet once again, this formal diversity goes along with a strong structural isomorphism. In particular, all languages display the same two morphological mergers on non-singular pronouns: merger of 1st inclusive and 3rd person (with a couple of exceptions) on the one hand, and merger of 1st exclusive and 2nd person on the other. This morphological pattern is specific to the three Vanikoro languages, and is not found in neighbouring Utupua (Tryon 1994:631). This may therefore constitute an important shared innovation diagnostic of a Vanikoro subgroup (see §1.3.2).[14]

Finally, our three languages add to the set of free pronouns another personal category, namely 3rd PLURAL INDEFINITE. This category has a special form as a free pronoun — TEA *idi* ~ LVN *nili* ~ TNM *deli*, which may be glossed 'people' (cf. French *on*). When this free pronoun is the subject, the agreement marker on the verb will be an ordinary 3rd plural prefix.

3.1.3 Possessive classes

The three languages also agree in the morphosyntax of possession. They all encode inalienable possession identically, by juxtaposing the possessed noun and its possessor. In the absence of possessive suffixes, inalienable possessors are encoded with the

[13] Instead of being coded by dedicated suffixes as in POc, objects and inalienable possessors are expressed by independent pronouns, which form distinct phonological words. The only exception to this principle is the Lovono suffix *-ŋo* for 2sg objects and possessors, which is distinct from the free pronoun *ago*, see Table 13 below.

[14] Pronoun systems often provide crucial diagnostic evidence in subgrouping research (Ross 2005).

independent personal pronoun: e.g. TEA *awa* ***ini*** ~ LVN *warene* ***ŋani*** ~ TNM *vasare* ***nini*** /throat 3sg:INDEP/ 'his throat'. Semantically, inalienable possession covers most body parts, plus a handful of intimate belongings, e.g. TEA *bete* ***ene*** ~ LVN *bele* ***ŋane*** ~ TNM *be* ***nana*** /mat 1sg:INDEP/ 'my bedmat'.

A possessive classifier is required for alienable types of possession. Four possessive categories can be distinguished (I indicate in square brackets the Teanu form of the classifier for 1sg possessor): FOOD [*enaka*]; DRINK [*me ene*]; KINSHIP [*one*]; GENERAL possession [*enone*], used as a default.

The category of FOOD possession is larger than its label suggests. First, it covers food (including items only chewed, like areca nut and betel leaf) as well as drink, thus overlapping with the dedicated DRINK classifier. Second, it is required for most tools ('knife', 'adze', 'spear', 'hook', 'box'…), plus the generic term 'belongings'. Third, it is used for 'language' and 'custom'.

(7)

TEA	*U-labu*	*ŋatene*	***enaka***	*u-lui*	*ne*	*mwoe*	***enone***.
LVN	*Ku-lo*	*vesemele*	***aŋa***	*ku-lawoi*	*ne*	*moe*	***iaŋa***.
TNM	*Go-nou*	*vamora*	***ae***	*go-lao*	*ini*	*nalama*	***ie***.
	2SG:IR-hold	thing	FOOD:1SG	2SG:IR-take.away	in	house	GEN:1SG

'Get my$_{[FOOD]}$ belongings and take them to my$_{[GENERAL]}$ house.'

(8)

TEA	*Dapa*	***iakapa***	*kape*	*le-mui*	*piene*	***akapa***.
LVN	*Detu*	***iegitore***	*gape*	*se-moi*	*mwamwane*	***agitore***.
TNM	*Dato*	***egeto***	*mota*	*la-muo*	*puiene*	***ageto***.
	PL:HUM	GEN:1INC:PL	FUT	3PL:IR-not.know	speech	FOOD:1INC:PL

'Our$_{[GENERAL]}$ people are going to forget our$_{[FOOD]}$ language.'

Once again, the three languages of Vanikoro agree perfectly on the semantic content of their formal categories.

3.1.4 Space directionals

Another domain where Teanu, Lovono and Tanema share identical structures is the system of space reference. Even though their adverbial directionals show impressive formal diversity (Table 9), their functional properties are parallel: they all resort to the 'in'–'out' contrast to encode the sea–land axis, and use the 'up'–'down' pair to encode a fixed cardinal axis, oriented towards southeast. While this system is attested elsewhere among Oceanic languages, it is distinct from the one reconstructed for POc (François 2004).

Table 9: Space directionals in their local and geocentric uses

Local use	Geocentric use	Teanu	Lovono	Tanema
'in'	'inland'	*takoie*	*ŋate*	*eto*
'out'	'seaward'	*tetake*	*mwaroa*	*emo*
'up'	'toward SE'	*tev' iu*	*ŋau*	*iu*
'down'	'toward NW'	*tev' tawo*	*lenu*	*ese*

3.2 Lexicon

The structural isomorphism so characteristic of Vanikoro languages relates not only to the morphosyntax, but also to the semantic organisation of the lexicon. I will mention successively two types of subdomain where this parallelism can be observed: the lexicon proper, and the phraseology.

3.2.1 Lexical semantics

We have seen that Teanu, Lovono and Tanema provide each grammatical category (possessive classifiers, space directionals …) with essentially the same semantic outline. The same can be said of lexical items and their meaning: when two quite distinct meanings are 'colexified' in one language — i.e. are expressed by the same lexical form (François 2008b) — the same pattern of colexification will almost certainly be found in the two other languages. Setting aside cases of polysemy which are shared by all or most Oceanic languages (e.g. 'hear'–'feel' …), some of the most distinctive examples of colexification are shown in Table 10.

Table 10: Most colexification patterns are shared across Vanikoro languages

Sense 1	Sense 2	Teanu	Lovono	Tanema
'one, single'	'the same'	*iune*	*tilu ~ tilioko*	*omwano*
Indefinite SG	'another'	*iote*	*leka*	*keo*
Indefinite PL	'others'	*kula*	*kule*	*kule*
'all'	'many'	*abia*	*maraŋa*	*abia*
'bird'	'friend'	*menuko*	*menuka*	*manuke*
'light (adj.)'	'dry'	*mimione*	*mimiane*	*mamiene*
'black'	'unripe'	*boro*	*bware*	*betika*

In the domain of compounding, a special case can be made regarding three adjectives, represented in Table 11. Despite their formal differences, they appear to be everywhere analysable in the same way, as if the result of calquing. Most forms are synchronically transparent, and based on the noun 'name'. As for TEA *ŋasune*, it can be analysed as historically a combination of *(e)ŋa* 'name' and *iune* 'one, the same'.

Table 11: Lexical connection between noun 'name' and three adjectives

English	Teanu	Lovono	Tanema	Literally
'name'	*eŋa*	*neŋe*	*niŋe*	
'identical'	*ŋasune*	*neŋe-tilu*	*niŋe-omwano*	'name-one'
'different'	*eŋa-iote*	*neŋe-leka*	*niŋe-keo*	'name-other'
'various'	*eŋa-eŋa*	*neŋe-neŋe*	*niŋe-niŋe*	'name-name'

Likewise, each of the three languages derives its intensifier 'very much, too much' — see ex.(5) — from its adjective 'bad'.[15] This connection is especially noteworthy as it involves distinct roots in each language (Table 12). The second element in these compound forms is obscure.

Table 12: Lexical connection between adjective 'bad' and intensifier

English	Teanu	Lovono	Tanema
'bad'	*tamwaliko*	*visale*	*vae*
Intensifier	*tamwa(liko)se*	*visale-wabeu*	*vae-pamabo*

3.2.2 Phraseology

The structural isomorphism between the three languages of Vanikoro is equally obvious from their phraseology, i.e. the routinised way in which they connect words together. I will only mention here a couple of original cases.

The three languages have an inalienable noun for 'body', to which they attach a variety of meanings, including 'genuine, true'[16] and 'beautiful', see Table 13.

Table 13: The polysemy of the noun 'body'

TEA	*ebele*	*eo*	*ebele*	*piene*	*ebele*	*kuo*	*ebel'*	*ini*
LVN	*nebele*	*-ŋo*	*nebele*	*mwamwane*	*nebele*	*nawe*	*nebele*	*ŋani*
TNM	*nibela*	*go*	*nibela*	*puiene*	*nibela*	*goia*	*nibela*	*nini*
	body	2sg	body	speech	body	canoe	body	3sg
	'your body'		'true words, truth'		'canoe hull' ~ 'beautiful canoe'		'his/her/its body' ~ 'Wonderful!'	

The neck or throat evidently constitutes, in Vanikoro, the seat of emotions and feelings.[17] The corresponding noun is found in a variety of formulas:

- 'I'm angry' is literally '*My throat is burning*'
- 'I'm sad' is literally '*My throat is blocked*'

Our three Vanikoro languages are strictly parallel in all these formulations. This is a fact of structural isomorphism, or calquing, as the words for 'throat' do not appear to be cognate across languages:

---

15 The connection is also attested in English (*I want it badly*), and closer to Vanikoro, in the Torres languages of Vanuatu: e.g. Lo-Toga *na luwō hia*, lit. 'it's big bad' = 'it's too big' (François, pers. data).

16 Interestingly, the languages of northern Vanuatu share the same colexification pattern between 'body' and 'true'. François (2005:501) thus proposes to reconstruct, for the common ancestor of north Vanuatu languages, a protoform *tur[i,u](ɣi) 'body, trunk; the real, main, very X; really'.

17 Osmond (2007) reports similar metaphors of emotions located in the larynx, for languages of the Southeast Solomons, as well as for the languages of the Trobriand Islands (after Malinowski 1922:408).

(9) TEA ***Awa*** *kupa* *i-su.*
LVN ***Warene*** *gamitu* *i-tu.*
TNM ***Vasare*** *gamuto* *i-to.*
throat 1EX:PL:INDEP 3SG:R-blocked
[lit. 'Our throats are blocked.'] 'We're sad ~ We're sorry.'

The sense 'like'/'want' is expressed by an unusual formula using a verb 'hit', taking the 'throat' as its subject:

– 'I like/want this' is literally '*My throat is hitting this.*'

The sentence becomes even more unusual when it is followed by an object clause ('want to do'), because it then involves a complementiser which is literally a verb meaning 'say'.[18] The subject of 'say' is normally the 'throat' itself (hence 3sg agreement), but occasionally it agrees syntactically with the throat's possessor:

– 'I want to [sleep]' is literally 'My throat is hitting IT SAYS I [sleep] …'
or 'My throat is hitting I SAY I [sleep] …'
which is often shortened to 'My throat I SAY I [sleep] …'

(10) TEA ***Awa*** *ene* *(i-viaene)* *ni-ko* *ne-mokoiu.*
LVN ***Warene*** *ŋane* *(i-piaine)* *ni-pu* *ka-mepeu.*
TNM ***Vasare*** *nana* *(i-vini)* *ni-po* *na-matou.*
throat 1SG:INDEP 3SG:R-hit 1SG:R-say 1SG:IR-sleep
'I want to sleep.'

In this case just as in all other contexts, the three languages can be translated literally, morpheme-by-morpheme, with no loss in idiomaticity or change in meaning. All one has to do is keep the structural — grammatical and lexical — boxes, and swap their phonological contents.

4 Addressing the paradox

In sum, the three languages of Vanikoro can be characterised by two contradictory properties. On the one hand, their fundamental genetic relatedness is blurred by a high degree of dissimilarity in the phonological forms of words, whether in the lexicon or in the morphology. But on the other hand, their grammatical categories and semantic structures show no equivalent to this formal diversity: instead, the three languages reveal perfect isomorphism, in each and every corner of their system. To paraphrase a formula by Sasse (see fn.19 below), they could ultimately be described as '*a single language with different vocabularies*'.

The question arises of what historical scenario would best explain this paradox, where divergence goes along with convergence. A simple explanation that comes to mind when accounting for the lexical diversification of cognate languages, might focus on the physical

[18] The grammaticalisation of a verb of saying into a complementiser is typologically common (Heine and Kuteva 2002; Chappell 2008). To take an Oceanic example, the verb 'say' in Araki, Vanuatu (François 2002), has exactly the same properties as in Vanikoro languages, including the persistance of a fully verbal morphology even when used as a complementiser.

separation between language communities. The absence, still today, of any land path relating villages across Vanikoro island, and the stories of ongoing fierce territorial fights between its three tribes, would then be understood as genuine evidence for geographical or social isolation, and thus as a possible key for the high degree of formal divergence between Teanu, Lovono and Tanema. However, several facts seem to contradict this diagnostic. First, the relatively small size of the island is at odds with the notion of a neat separation between the three tribes. And more crucially, their extreme degree of structural isomorphism is likely to reflect not only cases of shared retentions from a common ancestor, but also later linguistic convergence induced by language contact. In other words, the explanation resorting to the mere physical separation between communities does not tell the whole story.

The solution to the puzzle will probably have to be found not in the factual features of geography, but in the more subtle dimension of sociolinguistic behaviour. Indeed, a conspicuous characteristic of cultures in certain parts of Melanesia — in comparison, for example, with the Polynesian world (see Pawley 1981) — seems to be a social preference for small-scale social communities with no marked hierarchy between them, as well as a strong emphasis put on whichever anthropological or linguistic features may *differ* from one community to the other. Heterogeneity between villages or village groups tends to be socially valued as a way to construct a world of diversity, where each community is endowed with its own identity. In this framework, a local innovation in cultural and linguistic forms will tend to be perceived, and eventually retained, as emblematic of a specific group. Over time, this behaviour favours the emergence of cultural and linguistic divergence between erstwhile homogeneous communities. Interestingly, some language groups can be said to have only gone down this track to the point when the languages began to lose mutual intelligibility; but what is conspicuous in the case of Vanikoro languages, is that they seem to have pushed the process of differentiation far beyond that point, as though they were to keep diverging for ever.

In order to account for similar facts in other parts of Papua New Guinea, Thurston (1989), and later Ross (1996; 2001:155), have used the term ESOTEROGENY:

> *Esoterogeny* is a process that adds structural complexity to a language and makes it more efficient as a medium of communication among people of the same social group, while making it more difficult for outsiders to learn to speak well. (Thurston 1989)
>
> *Esoterogeny* arises through a group's desire for exclusiveness. (Ross 1996:184)
>
> If the members of a community have few ties with other communities and their emblematic lect is not usually known to outsiders, then they may use it as an 'in-group' code, an 'esoteric' lect from which outsiders are consciously excluded. Innovations leading to increased complexity and to differences from neighbouring lects will be favoured. (Ross 1997:239)

One could probably discuss the degree to which such sociolinguistic processes are 'conscious', and also how they interfere with motivations of various kinds (semantic, structural, pragmatic) in bringing about change. This being said, one can probably accept the general idea behind Thurston's concept, that language differentiation in Melanesia, far from being just an accident of geographical isolation, is largely influenced by a certain social attitude whereby each group tends to produce — whether consciously or not — its own distinctive speech tradition.

Now, while this hypothesis may help explain the high amount of lexical innovation and formal divergence that took place between Vanikoro languages, it seems at odds with the remarkable stability that we've observed among their structures. I would suggest this mismatch can be explained by the different nature of the linguistic components involved here. For one thing, the phonological form of the words (Saussure's 'signifiant', Grace's 'lexification'), whether lexical or grammatical, is the component most salient and conspicuous to the speakers' conscience, and therefore most likely to be preempted by motivations based on social emblematicity. Conversely, the structural and semantic dimension of language (Saussure's 'signifié', Grace's 'content form') would fall out of reach of the speakers' immediate linguistic awareness, in a way that would make it exempt of the sociolinguistic force of *esoterogeny*. Instead, structures tend to obey a totally contrary force, typical of language-contact situations,[19] that leads them to diffuse and converge: this is when multilingual speakers feel the pressure 'towards word-for-word translatable codes' (Gumperz 1971:270). The structural isomorphism that can be observed today among Vanikoro languages has the considerable advantage, for the bilingual speaker, of reducing any translation loss, thereby increasing the efficiency of cross-linguistic communication, and facilitating the cognitive processing of speech.

An important corollary of this whole reasoning is the necessity to distinguish two different components of language, because their evolution through history can follow quite distinct paths:

> The two components of language — content form and lexification — (…) evolve independently because (…) they are responding to different selective pressures, and those selective pressures are different because the functions of the two components are different. (…) It is the lexification on which the emblematic burden ultimately falls. (Grace 1981:30)

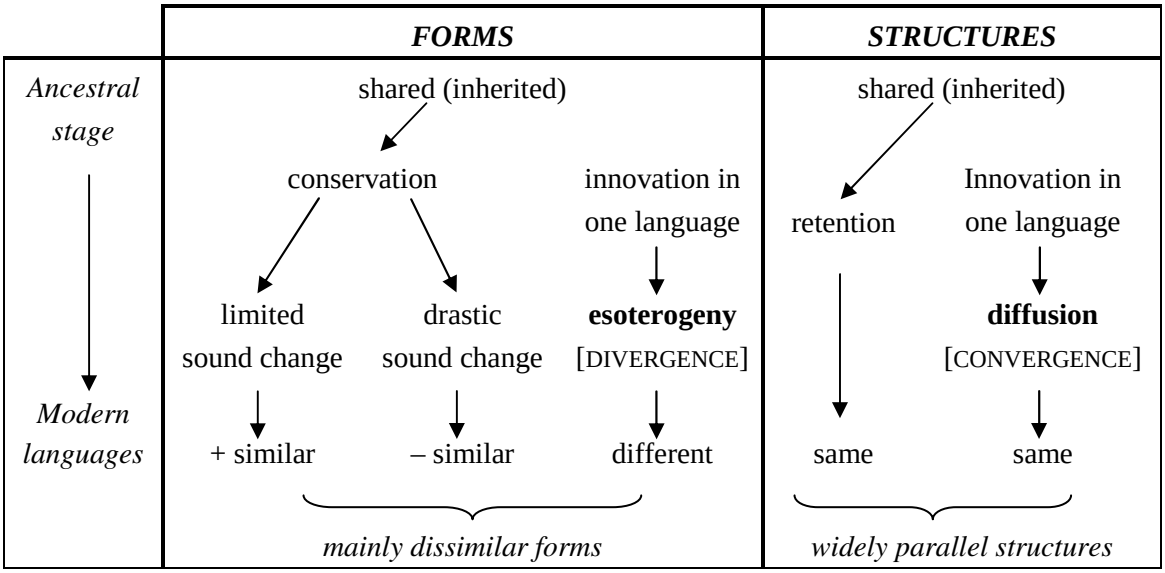
Vanikoro illustrates an extreme case of this possible mismatch between the two components of language.

Table 14 summarises a possible scenario about the sequence of developments in the evolution of the Vanikoro languages.

---

[19] Among many other references, see in particular Malcolm Ross' (1996, 1997, 2001) concept of *metatypy*, i.e. the typological alignment of one language to the structures of a neighbouring language, through linguistic contact. Ross (2001:149) also cites this statement by Sasse (1985): 'With advanced language contact, there arises the tendency to develop a single language with different vocabularies.'

Table 14: Different forms, shared structures among Vanikoro languages: a historical scenario



5 Conclusion

The comparison of Teanu, Lovono and Tanema reveals the intricacies of the island's local history. The strong isomorphism found between the structures of these languages betrays their remote common ancestry, as much as it points to a history of intense language contact which the three tribes, *nolens volens*, have lived through over the centuries. On the other hand, the actual word forms found in their vocabularies and morphology have tended to follow a powerful tendency towards diversification, in accordance with the speakers' tacit perceptions that the three communities, often caught in conflict and territorial hostilities, should sound and feel to be distinct social groups.

Overall, the paradox observed among the three modern languages of Vanikoro — *dissimilar forms, similar structures* — results from the interplay between these two contradictory forces: a socially driven push to increase language differences versus a functionally grounded tendency to minimise them.

References

Camden, William. 1979. Parallels in structure and lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. In *Papers in Pidgin and Creole linguistics, No.2*, 51–117. Canberra: Pacific Linguistics.

Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12.45–98.

Dillon, Peter. 1830. *Voyage aux îles de la mer du Sud, en 1827 et 1828, et Relation de la découverte du sort de La Pérouse. Par le Capitaine Peter Dillon*. Paris: Pillet Aîné.

Dumont d'Urville, Jules Sébastien César. 1830–1834. *Voyage de la corvette l'Astrolabe, exécuté par ordre du Roi, pendant les années 1826–1827–1828–1829, sous le commandement de M.J. Dumont d'Urville, Capitaine de vaisseau*. 23 volumes. Paris: J. Tastu.

Dyen, Isidore, Joseph B. Kruskal, and Paul Black. 1992. *An Indoeuropean classification: a lexicostatistical experiment*. Transactions of the American Philosophical Society, vol. 82(5). Philadelphia: The American Philosophical Society.

Enfield, Nicholas. 2001. On genetic and areal linguistics in mainland Southeast Asia: parallel polyfunctionality of 'acquire'. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 255–290. Oxford: Oxford University Press.

Firth, Raymond. 1985. *Tikopia–English dictionary — Taranga fakatikopia ma taranga fakainglisi*. Auckland: Auckland University Press.

François, Alexandre. 2002. *Araki. A disappearing language of Vanuatu*. Canberra: Pacific Linguistics.

—— 2004. Reconstructing the geocentric system of Proto Oceanic. *Oceanic Linguistics* 439(1).1–32.

—— 2005. Unraveling the history of the vowels of seventeen northern Vanuatu languages. *Oceanic Linguistics* 449(2).443–504.

—— 2006. Are Vanikoro languages really Austronesian? Paper presented at the Second conference on Austronesian Languages and Linguistics. Oxford, UK.

—— 2007. Parallel meanings, divergent forms in the North Vanuatu Sprachbund: diffusion or genetic inheritance? Paper presented at 7th International Congress of the Association of Linguistic Typology. Paris, France.

—— 2008a. Mystère des langues, magie des légendes. In Association Salomon, ed. *Le mystère Lapérouse ou le rêve inachevé d'un roi*, 230–233. Paris: de Conti – Musée national de la Marine.

—— 2008b. Semantic maps and the typology of colexification: intertwining polysemous networks across languages. In M. Vanhove, ed. *From polysemy to semantic change: towards a typology of lexical semantic associations*, 163–215. Amsterdam/ Philadelphia: John Benjamins.

—— forthcoming. Des valeurs en héritage: les isomorphismes sémantiques et la reconstruction des langues. In I. Choi-Jonin, M. Duval and O. Soutet, eds *Typologie et comparatisme. Hommages offerts à Alain Lemaréchal*. Orbis-Supplementa 29. Louvain: Peeters.

—— in preparation. Language ecology and change in north Vanuatu: a tale of divergence and convergence.

Gaimard, Joseph Paul. 1833. Extraits du Journal de M. Gaimard. In J. Dumont d'Urville, *Voyage de la corvette l'Astrolabe, exécuté par ordre du Roi, pendant les années 1826–1827–1828–1829, sous le commandement de M.J. Dumont d'Urville, Capitaine de vaisseau — Histoire du voyage*, vol. 5, 322–353. Paris: J. Tastu, éditeur.

—— 1834. Vocabulaires des idiomes des habitans de Vanikoro. In J. Dumont d'Urville, *Voyage de découvertes de l'Astrolabe, exécuté par ordre du Roi, pendant les années 1826-1827-1828-1829, sous le commandement de M.J. Dumont d'Urville, Capitaine de vaisseau — Philologie*, vol. 1, 165–174. Paris: Ministère de la Marine.

Grace, George. 1981. *An essay on language*. Columbia, SC: Hornbeam Press.

—— 1990. The 'aberrant' (vs. 'exemplary') Melanesian languages. In Philip Baldi, ed. *Linguistic change and reconstruction methodology*, vol. 45, 155-173. Berlin: Mouton de Gruyter.

Gumperz, John. 1971. Convergence and creolization: a case from the Indo-Aryan/ Dravidian border in India. In *Language in social groups: essays by John J. Gumperz*, 251–273. Stanford: Stanford University Press.

Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Hjelmslev, Louis. 1961 [1943]. *Prolegomena to a theory of language*. Madison: University of Wisconsin Press.

Ivens, Walter G. 1918. *Dictionary and grammar of the language of Sa'a and Ulawa, Solomon Islands*. Washington: Carnegie Institution of Washington.

Keesing, Roger M. 1988. *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.

—— 1991. Substrates, calquing and grammaticalization in Melanesian Pidgin. In E.C. Traugott and B. Heine, eds *Approaches to grammaticalization*, vol. 1, 315–342. Amsterdam/Philadelphia: John Benjamins.

Lynch, John. 2001. *The linguistic history of Southern Vanuatu*. Canberra: Pacific Linguistics.

Malinowski, Bronislaw, 1922. *Argonauts of the western Pacific*. London: Routledge and Kegan Paul.

Osmond, Meredith. 2007. Describing emotions and enduring dispositions in Oceanic languages. Paper presented at 7th International Conference on Oceanic Linguistics. Nouméa, New Caledonia.

Pawley, Andrew. 1981. Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In J. Hollyman and A. Pawley, eds *Essays in Pacific languages and cultures in honour of Bruce Biggs*, 269–310. Auckland: Linguistic Society of New Zealand.

—— 2006. Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 115.213–256.

—— 2007. Was there early Lapita settlement of the Solomon Islands? Bringing linguistic evidence to an archaeological debate. Paper presented at 7th International Conference on Oceanic Linguistics. Nouméa, New Caledonia.

Ross, Malcolm D. 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 Mark Durie and Malcolm Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 180–217. Oxford: Oxford University Press.

—— 1997. Social networks and kinds of speech-community event. In Roger Blench and Matthew Spriggs, eds *Archaeology and language 1: theoretical and methodological orientations*, 209–261. London: Routledge.

—— 2001. Contact-induced change in Oceanic languages in north-west Melanesia. In Alexandra Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 134–166. Oxford: Oxford University Press.

—— 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In A. Pawley, R. Attenborough, R. Hide and J. Golson, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–66. Canberra: Pacific Linguistics.

Ross, Malcolm D. and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.456–498.

Sasse, Hans-Jürgen. 1985. Sprachkontakt und Sprachwandel: Die Gräzisierung der albanischen Mundarten Griechenlands. *Papiere zur Linguistik* 32.37–95.

Siegel, Jeff. 2008. *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.

Thurston, William. 1989. How exoteric languages build a lexicon: esoterogeny in western New Britain. In Ray Harlow and Robin Hooper, eds *VICAL 1: Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics, part 2*, 555–579. Auckland: Linguistic Society of New Zealand.

—— 1994. Renovation and innovation in the languages of north-western New Britain. In T. Dutton and D. Tryon, eds *Language contact and change in the Austronesian world*, 573–609. Berlin: Mouton de Gruyter.

Tryon, Darrell. 1994. Language contact and contact-induced language change in the eastern outer islands, Solomon Islands. In T. Dutton and D. Tryon, eds *Language contact and change in the Austronesian world*, 611–648. Berlin: Mouton de Gruyter.

—— 2002. Buma. In J. Lynch, M. Ross and T. Crowley, eds *The Oceanic languages*, 573–586. London: Curzon.

Tryon, Darrell and Brian D. Hackman. 1983. *Solomon Islands languages: an internal classification*. Canberra: Pacific Linguistics.

Tryon, Darrell, ed. 1995. *Comparative Austronesian dictionary. An introduction to Austronesian studies*. Berlin/New York: Mouton de Gruyter.

Tua, Benjamin and Peter Lincoln. 1979. *Our custom way Vanikoro: the custom book*. Auckland: University of Auckland.

Wurm, Stephen A. 1976. The Reef Islands–Santa Cruz family. In S.A. Wurm, ed. *New Guinea area languages and language study*, vol. 2: *Austronesian languages*, 637–674. Canberra: Pacific Linguistics.

7 *Expanding character sets for phylogeny: a Slavic test case*

JOHANNA NICHOLS

1 Introduction

Linguistic phylogeny, computational and otherwise, is only as good as the philological work behind the items it compares, and one of the most durably useful contributions a historical linguist can make is a large data set for a language family with reliable data accurately coded for use in phylogenetic and comparative work. This paper is a first step toward compiling such a data set for the Slavic languages. It was inspired by Ross (1996, 1997, and especially 1988), where extensive and carefully assessed comparative-historical data made it possible to solve long-standing problems of phylogeny and reconstruct branch splits and contact histories.

The Slavic family is a good test case for phylogeny because there is good reason to believe the received family tree with its three primary branches (East, West, and South Slavic) does not straightforwardly reflect the history of dispersal: South Slavic originated not in a unitary ancestral dialect but from a subsequent fusion of what were initially separate branches; literary standard languages reflect a certain amount of dialect fusion and leveling; and what can be known about the pre-dispersal dialectology points to a very different subgrouping (e.g. Andersen 1996, 1969; Trubachev 1967). The Slavic languages are a workable test case because the phonological and morphological changes in the family are well understood and there are excellent synchronic and etymological dictionaries.

This paper examines the posture verbs ‘sit’, ‘stand’, and ‘lie’ in Slavic languages. Stance verbs are generally considered stable vocabulary, and the glosses ‘sit’, ‘stand’, and ‘lie’ appear in the Swadesh 100-word list. However, finding matches for these simple glosses is problematic for many languages, Slavic included. Each Slavic language distinguishes a static verb (e.g. ‘sit, be sitting, be in sitting position’) from a dynamic one (‘sit down, assume sitting position’) and a transitive (‘seat, have sit, let sit, put in sitting position’), and also distinguishes perfective from imperfective dynamic and transitive verbs, for a total of five lexemes per posture type. The set of aspect and Aktionsart forms for each posture are generally derived from the same Indo-European root by means of affixation, ablaut alternations, and conjugation class alternations. The morphological forms and classes associated with these derivations and the properties of their various stems can be used to extract a total of 8 forms x 5 lexemes x 3 stances = 120 independent characters

in place of the three Swadesh items. This paper shows how this expanded set of characters can be exploited in phylogeny and comparison. The Appendix presents full data sets for the static and dynamic intransitive forms (8 x 3 x 3 = 72 characters), coded for character states, in the modern Slavic languages plus ancestral Common Slavic, with discussion of the choice of forms and their encoding.

When morphological characters are used in phylogeny, they usually figure as individual morphemes (e.g. the presence of the Indo-European verb augment in Greek, Armenian, and vestigially in Indo-Iranian), paradigms (e.g. the mobile-stress noun paradigms of Balto-Slavic), or innovations (e.g. number syncretism in third person verb desinences in Baltic; loss of all but the *o-/a-stem adjective declensions in Slavic). Here they are used differently, basically as segments of individual wordlist items: does the imperfective dynamic verb 'sit' have a prefix? what is its present-stem root vocalism? its conjugation suffix? is it reflexive? etc. These are independent options available in word formation, so a given sequence of morphological character states in a given word is reasonably arbitrary relative to the meaning of the word. Characters used in this way, for morphological segmentation, are not ordinary morphological characters and deserve a distinct term (I do not suggest one here). They could be used (as individual morphological characters could not) for calculating spelling or edit distances between words (Johnson 2008:201–208; Nichols and Warnow 2008:795–796, both with references; these are also known as Levenshtein distances and have to do with how many steps would be required to get from one word to another, segment by segment). They could be used instead of or in addition to distances calculated on phonemes or letters.

I also give some simple tabulations of character sharings for each pair of Slavic languages and some simple calculations of numbers of sharings and extent of affinity between the modern languages. The recognised phylogenetic branches and contact situations are apparent in the results to some extent, but there are also some connections that are unexpected. This tells us that morphological characters derived from segmenting wordlist items can add information to family trees.

My main purpose is not to revise the Slavic family tree but to illustrate a method for breaking cognate words down into additional components to create a larger character set for distance measures and phylogeny. Phylogenetic work in linguistics often uses character sets that are large by linguistic standards. Ross 1988 drew on about 350 Proto Oceanic cognate sets plus additional items from lower branches plus a number of grammatical constructions and paradigms. Nakhleh et al. (2005) have 294 characters from 24 Indo-European languages (p.392; also Ringe et al. 2002). A great deal of philological labour went into compiling those character sets, and they more or less exhausted the available cognate sets (Ross) and diagnostic innovations (Nakhleh et al.), yet they are one or possibly two orders of magnitude fewer than the number of characters needed to construct a solid phylogeny using distance measures (Nichols and Warnow 2008:784), which in computational biological phylogeny may require tens of thousands of characters to find a good tree — and distance measures are some of the most promising and versatile for subgrouping, especially in language families where reconstruction is not far along. One reason why so many characters are needed is that biological evolution presents many cases of independent parallel evolution, back-mutation (a mutation that happens to restore the original DNA sequence and the original phenotype), and other developments that obscure the actual evolutionary history. In language change these are less rampant and more detectable, so fewer characters may suffice. Still, various techniques of computational

phylogeny proceed by generating the very large number of trees required to capture all characters (as an example, if one innovation is shared by languages *a* and *b* and another by *b* and *c*, two trees can be drawn: one with branch *ab* and the other with branch *bc*), scanning those trees and choosing the best ones (those with the fewest changes overall), after which the linguist examines the best trees and rates them on linguistic criteria. The larger the number of characters, the less the impact of any one change that crosscuts branches and the more robustly the actual evolutionary history can emerge. Thus, for language families with considerable time depth and divergence (such as Indo-European and Austronesian), several thousand characters may be needed for firm identification of a good phylogeny. This is a problem, as the total number of usable lexical and grammatical roots, morphological forms and paradigms, and sound changes that can be found in even a well-described language family would probably number under one thousand.

Therefore this paper suggests a way to squeeze some additional characters out of cognate sets. It is a programmatic account, with only a tiny sample of data. The method is preliminary, and it will work only in languages with some overt derivational and inflectional morphology. It requires good reconstructions and precise grammatical analysis of the daughter languages. Like most kinds of linguistic phylogenetic characters, these offer no protection against mistaking independent parallel development for reconstructability. But it does increase the number of characters, and has the possibility of improving phylogeny and distance measures. Since for an old and divergent family it is impossible for a linguist to construct and compare all possible trees for so many characters, it must be done computationally. The great benefit of all this additional work is that it offers the prospect of definitively resolving questions of tree structure and language family evolution.

2 Data and method

Choosing a single match for the Swadesh wordlist items ‘sit’, ‘stand’, and ‘lie’ is problematic for languages that have more than one word for one or more of these glosses. Slavic languages generally have three, illustrated for ‘sit’ in Table 1.

Table 1: Slavic verbs for ‘sit’

	Russian		Czech	
	infinitive	present (3SG)	infinitive	present (1SG)
Static (‘sit, be sitting’)	*sidet’*	*sidit*	*sedět*	*sedím*
Dynamic, perfective (‘sit down’)	*sest’*	*sjadet*	*sednout si*	*sednu si;*
		posadit se		*posadím*
Dynamic, imperfective	*sadit’sja*	*saditsja*	*sedat si*	*sedám si;*
	posazovat se	*posazuji se*		

Slavic languages distinguish perfective vs. imperfective forms of most verbs. The static verb is imperfective and means ‘be in a sitting position’. The perfective dynamic means ‘sit down (once, or on a particular occasion)’. The imperfective dynamic means ‘sit down’ (more than once — iterative, habitual, etc.)’ and/or ‘be in the process of sitting down’. (For the meanings of perfective and imperfective in various Slavic languages see Dickey 2000.) These three aspect/Aktionsart classes make up what I will call the triad for posture verbs in all modern Slavic languages. The three verbs in the triad are lexically basic (appearing e.g.

in first-year language textbooks), but there are other Aktionsart derivatives of the same roots. Most of the Slavic languages also form a delimitative static verb meaning 'sit for a short time' and/or 'sit for a specified time' (e.g. Russian *posidet'*), some have iteratives of this and/or the dynamic forms, and some have culminatives, resultatives, totalising forms, and others (meanings such as 'sit through to the end', 'sit (someplace) too long, to excess', etc.) (for discussions in English of some of these forms see Timberlake 2004:403–406; Dickey 2000:8–9; Townsend 1975:118ff.; Forsyth 1970:20–26). There are also transitive forms ('seat, have sit, put in sitting position', etc.) and their imperfectives, which are not covered here.

Most Slavic verbs have two stem forms: one used in the present or nonpast tense and one used in the infinitive, aorist, or past forms. I will call them *past* stem and *present* stem, though these labels gloss over various morphological sticking points. The forms in Table 1 illustrate the two stems of Russian and Czech. The two stem forms differ from each other in various ways, and different conjugation classes have different endings. Some of the forms have reflexive morphology (Russian *-sja*, Czech *si/se).* Czech has two synonymous forms for the dynamic, differently derived with reflexivisation and a prefix. Despite these differences, all the forms from both languages have the same root *sVd-*, the Indo-European root for 'sit'.

The differences in stem form, conjugation class, etc. can be reduced to the following questions:

- What is the root?
- What is the vocalism of the past stem?
- What is the vocalism of the present stem?
- What is the conjugation suffix (if any) of the past stem?
- What is the conjugation suffix (if any) of the present stem?
- What is the aspect?
- What is the prefix (if any?)
- Is the verb reflexive?

Coding for these questions yields a componential analysis that gives eight separate features for use in comparison, instead of a single unanalysed word. For the triad of forms shown in Table 1 there are a total of 24 features; the triads for three postures yield a total of 72. I surveyed these three forms of posture verbs in nearly all of the modern Slavic languages plus the reconstructed Late Proto Slavic system to test the usefulness for phylogeny and distance measures of this larger character set. The Appendix gives a past-stem and a present-stem form for each verb in each language, together with a character state coding showing which languages have the same form in this or that cell of their paradigms. As an example of character coding, in the Russian and Czech examples in Table 1, the only shared characters in the dynamic Russian *sest'*, Czech *sednout si* are the root vocalism in the present stem and reflexivisation in the imperfective dynamic. But Slovenian dynamic *sesti* has all of its character states identical to Russian, and in the static present forms Polish *siedzi* and Bosnian/Croatian/Serbian (BCS) *sjedi* have all character states identical to Czech.[1]

[1] Slavic verb paradigms also have the additional properties of stress type (root versus post-root, fixed versus mobile) and conjugation class (e.g. thematic versus athematic), but these are not included here because the changes they have undergone have generally been across-the-board ones instead of individual ones differentiating lexemes: West Slavic languages have developed fixed stress; several languages generalise the athematic conjugation by redistributing it across suffix types.

The main intended contributions of this paper are the Appendix and the discussion (in the next section) of historical questions, especially decisions as to whether the various identical character codings reflect shared innovations, shared retentions, or independent parallel developments. I hope the data set will be useful to phylogeny and diffusion studies for a long time, and in any case it will cause Malcolm Ross's name to be cited in Slavistic publications more than it now is.

3 Historical issues

Late Proto Slavic (LPS; also often known as Common Slavic) as a linguistic system existed from about the fifth century CE, when the Slavic dispersal began and Slavic speakers first came into contact with post-Roman Europe, and the ninth century, the date of the first written records and the time frame when at least the Old Novgorod dialect had ceased to be mutually intelligible with the rest of Slavic. Since some dialect differences antedate even the beginning of LPS (e.g. Andersen 1996), it seems that LPS was never a homogeneous variety. Still, it was a single extended speech community as indicated by lexical borrowings and sound changes that spread from one end of the range to the other.

East Slavic:		† Old Novgorod
		Russian
		Belarusian
		Ukrainian
West Slavic:	Lechitic	Polish
		Cashubian
		† Polabian
	Sorbian	Upper Sorbian
		Lower Sorbian
	Czechoslovak	Czech
		Slovak
South Slavic:	Eastern	Bulgarian
		Macedonian
	Western	Bosnian/Croatian/Serbian
		Slovenian

Figure 1: The Slavic family tree

The Slavic family tree is shown in Figure 1. It is based on modern phonological and some morphological sharings but is known to conflict with the actual dispersal and settlement patterns. The South Slavic area (the Balkan peninsula) was settled from both northeast and northwest, and to this day there are phonological and morphological east/west and north/south distributions reflecting that settlement pattern and not coinciding exactly with any language boundary. The only clear uniquely South Slavic phonological innovations — the merger of nasal *ę with plain *e and the early merger of *y and *i — occurred some centuries after the first writing and after divergences had begun to appear

within South Slavic. In East Slavic, in medieval times there was a four-way major dialect division, now effaced by modern boundaries and the spread of Russian influence. In medieval times the Old Novgorod dialect was very divergent and seems to have been the first Slavic language to lose mutual intelligibility with the others (Nichols 1993). The modern dialect differentiation between eastern and western Ukrainian is considerable (to the extent that westernmost Ukrainian, or Rusyn/Ruthenian, could almost be considered a separate language.) Slovenian, despite its small size, also has considerable internal dialect divergence.

The earliest written Slavic language, Old Church Slavic (OCS), does not fit into any one branch but is a written tradition comprising early West and South Slavic. The languages not covered here are Polabian, Kashubian, Lower Sorbian, Rusyn, Old Novgorod, and OCS.

Verb derivation in LPS utilised changes in ablaut, suffix, and/or conjugation class, most of them inherited from Indo-European, to bring about various differences in temporality and valence, as well as prefixes (many of them cognate to prepositions and/or adverbs, and as a class functionally similar and in some instances cognate to English verb particles, other Germanic prefixes, etc.). In LPS the considerable variety of possible verb lexemes gradually settled into perfective-imperfective aspect pairings and the triad of posture verb forms, and also served to build up the pairing of past and present stems. The main LPS forms of 'sit' were as shown in Table 2.

Table 2: LPS 'sit' in past and present stems

se:d-e:-	se:d-i:-	static	*(sěděti, sěd´ǫ)*
se:d-	se^{n}d-	dynamic	(*sěsti, sędǫ*)
se:d-a:-	se:d-j-/se:d-aj-	iterative	(*sědati, sěd'ǫ/sědajǫ)*
sed-no^{n}-	sed-n-	inchoative	(*sednǫti, sednǫ*)
sa:d-i:-	sa:d-i:-	transitive	(*saditi, sažu*)

Italicised forms are Slavistic transcription of the infinitive and first person singular present.[2]

Posture verbs are generally thought of as genealogically fairly stable, and indeed all Slavic languages preserve the Indo-European roots for these verbs. Nonetheless, apart from the shared root, changes in derivational form are fairly numerous. Some of these are early and reflect the creation of the triad, in particular the recruitment of an imperfective partner to the dynamic verb. Iterative morphology was commonly recruited to form imperfectives, and inchoatives sometimes displaced or influenced the dynamic forms (Dickey 2003). Most of the changes, however, seem to be individual levelings, adjustments, etc. in an already structured system. A behind-the-scenes force in the intransitive sets surveyed here is the transitive forms 'have (someone) sit/stand/lie', which are a source of changes in intransitives in many of the cells. This pressure shows up here in the reflexive forms (reflexivisation being the usual form of detransitivisation in Slavic), but there are other sets in which the vocalism of the intransitive has been adjusted to match that of the transitive without actual derivation from the transitive (e.g. Slovak dynamic *sadnut (si)* 'sit down').

[2] Long vowels in 'sit' are the result of Winter's Law, which lengthened vowels in Balto-Slavic before plain voiceless stops (the rule was blocked by a sonorant adjacent to the stop, as in the inchoative forms here) (Winter 1978; Kortlandt 1988).

Note that, in all three posture sets, the dynamic form is unsuffixed and its past stem, which has no present-tense nasal infix, is equal to the root in both Slavic and Indo-European.

The Appendix shows an LPS reconstruction for each stem of each verb form, based on standard sources (chiefly Vaillant 1966; Sławski 1952ff.; Trubachev, ed. 1974ff.: vols 14–16). Though it is not a foregone conclusion that the imperfective dynamic set had been incorporated into a fully crystallised triad in LPS times, I include the most likely candidate for that position.[3] The root, stem vocalisms, suffixes, and other properties for those LPS forms and the modern forms are shown together with the character state encodings: state 1 where the entry is the same as the LPS one, states 2 and higher for others (each morphologically different entry receives a different state coding). In any column, any two forms with the same character state number are the same (by dint of either common retention or common innovation). Entries of forms or stems are in modern orthography or transliteration. Entries for vocalism and suffixes are not the modern forms but the LPS forms they reflect. Where a modern phoneme is a merger of LPS phonemes (e.g. Bulgarian /e/ reflects LPS *e and unstressed *e:) I have assumed the regular LPS form: thus, for Bulgarian, long *e: vocalism in the first syllable of static *sedja*, short *e in dynamic *sedna).* The entry for the form is flagged by '#' where there has been merger, but the character state codings assume the regular antecedent.

The kinds of changes that have occurred are the following. Most drastic, an entirely different verb root now appears in one cell: Polish *kłaść się* 'lie' (dynamic imperfective) instead of a verb based on *leg- 'lie'. What has actually happened is that Polish transitive *położyć* 'lay; put' has a suppletive imperfective *kłaść,* and the dynamic intransitive forms (both perfective and imperfective) are derived by reflexivisation from the transitive. For the perfective this entails a change of root vocalism (transitive *położyć* has the root *log- while the LPS dynamic intransitive had *leg-) as well as reflexivisation and addition of the prefix *po-*. In the suppletive imperfective, the verb *kłaść się* comes with its own vocalism, conjugation type, and other properties, so the number of character state differences mushrooms if they are all counted (as I have done in the preliminary counts described below).

---

[3] Of the dynamic imperfectives, prefixed *vъstajati* 'stand up' has several attestations in the OCS canon (Sadnik and Aitzetmüller 1955). **Sědati* 'sit' occurs only in reflexive *sědati sę* 'collapse, shatter' and prefixed *sъsědati sę* 'curdle, coagulate; go numb'. *Lěgati* 'lie' has only one attestation, but that one is telling:

> *Mol<itva>* ***lěžǫšte*** *na loži*
> prayer.NOM lying.ADV on couch
> 'bedtime prayer', 'prayer while going to bed/lying down in bed'
> (Frček 1933:734 'Prière à dire quand on se couche')

The sense is not iterative but progressive or durative and describes the process of getting ready for bed or getting into bed. The short prayer ends in

> *sьde azъ, poklaněvъ sę* ***lešti*** *xoštǫ vъ imę O<tъ>ca ...*
> here I having.worshipped lie.DYN.INF want.1sg in name father.GEN
> And here I have worshipped and am going to/want to lie down [=go to bed] in the name of the Father ...

where the perfective dynamic form has clear perfective sense, referring to an event that is about to occur and be completed. Thus the two verbs are certainly functioning as a pair and the former iterative is now just an imperfective.

LPS inchoatives have supplanted the simple dynamic forms for ‘sit’ and ‘lie’ in Upper Sorbian, Czech, Slovak, Bulgarian, and Macedonian, and in the present stem in BCS. ‘Stand’ had inchoative morphology in its present stem in LPS, and this has spread to the past stem in Upper Sorbian, Macedonian, and Bulgarian. Macedonian has added an additional iterative suffix *-uva-* to the *-n-* suffix of the inchoative in its imperfective dynamic forms.

In ‘sit’ Czech has kept the inchoative form *sednout si* side by side with reflexive (i.e. detransitive) *posadit se* in dynamic ‘stand’, and in the imperfective the LPS iterative is retained together with a secondary iterative *posazovat se* of the reflexive.[4] The two verbs are equally basic according to speakers I have consulted, and both are found in dictionaries. This is a case of what is known in biological phylogeny as polymorphy: two distinct characters in one position or cell. Coding the character states is problematic in such cases, as is discussed below.

In dynamic ‘sit’ and ‘lie’ the vocalism of the present stem (containing the reflex of a nasal vowel) has extended to the past stem in Polish and Slovak, and in Ukrainian in ‘lie’ only.[5]

Pure sound change has produced what look at first glance like morphological changes in several forms. Czech *stát* ‘stand’ (static) and similar forms in Polish, Slovak, Upper Sorbian, and Slovenian may look like extensions of the vocalism of the dynamic to the static, but in fact they are the result of regular loss of intervocalic *j and coalescence of the two vowels. In all of East and West Slavic and Bulgarian the *-j-* has been replaced by *-v-* (a stem-final consonant in several high-frequency verbs, which has spread to become a productive iterative or secondary imperfective formative); this is a morphological change.

The only character that has never changed is aspect. Given the differences in the meanings and behaviour of aspect in the modern Slavic languages (Dickey 2000) and the demonstrably late development of modern aspect in Russian (Bermel 1997), it is clear that aspect has not been semantically stable for the past millennium. What is quite stable is the positions of verbs in aspect pairings: almost never do morphologically identical verbs have different aspects in different Slavic languages.

4 Preliminary findings

While doing an actual phylogeny would require many more characters, even from just these 72 one can draw some conclusions that indicate that this kind of breakdown is useful.

(a) *An overall index of lexical and morphological conservatism per language*. Table 3 shows the number of character states that each language shares with LPS. Consistent with the fact that Slavic languages are morphologically conservative overall, there is rather little spread and no obvious geographic or genealogical pattern to what differences exist.

---

[4] Forms of the reflexive clitic are accusative *se*, dative *si*. The accusative derives intransitives from transitives, and the dative makes no change in valence.

[5] The Ukrainian past tense forms are masc. *lih*, fem. *l’ahla. Lih* has not undergone the extension. Its vowel has been raised from original *e by the regular change of compensatory lengthening (Timberlake 1983a, b), and this change must have been at least underway when the morphological extension of vocalism occurred, protecting it from the extension (which must have taken the form of replacing the citation or most basic vocalism of the past stem, rather than the past-stem vowel per se, by the present-stem vowel).

Table 3: Total Late Proto Slavic retentions per language.
The logically possible total is 72.

Slovenian	65
BCS[a]	65
Ukrainian	63
Bulgarian	63
Polish	60
Belarusian	59
Macedonian	59
Czech	58
Russian	57
USorbian	55
Slovak	56

[a] BCS = Bosnian/Croatian/Serbian.

(b) *A measure of geographical or genealogical connectedness*. Table 4 shows the number of shared innovations in each language. Czech, Slovak, and Upper Sorbian share a good many innovations with various languages and each other, as might be expected of languages that are close sisters, immediate neighbors, and centrally located. Languages with few shared innovations are among the most peripherally located, as again seems logical. The two Slavic languages that belong to the Balkan Sprachbund, Bulgarian and Macedonian, are not equally prone and not particularly prone to share innovations with other languages, and this indicates that morphological differences of the kinds studied here may be resistant to areal convergence.

(c) *Structurally nearest languages*. Table 5 shows, for each language, the language(s) with which it has the most shared innovations, the number of those innovations, and whether the languages are in the same branch of Slavic. Two sets of very close sister languages — Russian and Belarusian, and Upper Sorbian, Czech, and Slovak — have the highest numbers of shared innovations, and nearly all languages (9 out of 11) have at least one closest connection within the branch, suggesting that the characters may capture subgrouping rather well. On the other hand, within-branch and extra-branch closest connections have similar frequency overall, and the southern, peripheral, Balkan languages Macedonian and Bulgarian are unexpectedly close to the northwestern, peripheral, and isolated Upper Sorbian, affinities that find no basis in anything we know about the history of contact in early Slavic.

Table 4: Total shared innovations per language

Czech	66
Slovak	64
USorbian	57
Belarusian	49
Bulgarian	43
Russian	43
BCS	39
Ukrainian	37
Polish	36
Slovenian	33
Macedonian	27

Table 5: Closest neighbour(s) for each Slavic language, and number of shared innovations

Russian	Belarusian*	12
Belarusian	Russian*	12
Ukrainian	Russian*, Slovak, Slovenian	5
Polish	Belarusian, Slovak*, Czech	5
Upper Sorbian	Czech*	13
Czech	Upper Sorbian*	13
Slovak	Czech*	12
Slovenian	BCS*, Ukrainian	5
BCS	Czech, Slovak	6
Bulgarian	Upper Sorbian	9
Macedonian	Upper Sorbian, Bulgarian*	6

* = language in the same major branch (East, West, South)

(d) *Typological conclusions*. The three posture verbs do not evolve together as a set; they undergo changes independently and are not equally prone to change. (1)–(2) show two hierarchies for overall proneness to morphological change among these posture verbs (for the figures see Table 6).

(1) 'lie' > 'sit' > 'stand'

(2) dynamic imperfective > dynamic perfective > static

Table 6: Number of innovative character states for the various verb forms

	stat	dyn	dyn impf	Total
Sit	5	15	20	40
Stand	2	11	20	33
Lie	0	22	38	60
Total	7	48	78	

Only in the verb that is most change-prone on both of these hierarchies, namely the dynamic imperfective of 'lie', is there wholescale lexical replacement producing suppletion (in Polish *kłaść się,* discussed above). On the other hand, this suppletion is inherited from the transitive forms from which the dynamic forms are derived (and from which intransitives tend to be derived throughout the vocabularies of all the languages). This is part of a larger pattern in which transitive members of paradigms undergo lexical renewal and thereby show that they occupy the semantic core of the set (in the mechanism of renewal described by Kuryłowicz (1947, 1964:11–15); Nichols (2006) argues this for posture verbs in other languages as well). Though the perfective of the dynamic verb is often the morphologically simplest of its triad, and was clearly so in LPS, it is the static form that has been most resistant to morphological change. Curiously, though 'lie' is prone to morphological change overall, its static form has had no morphological changes.

5 Discussion and problems

I believe this effort has shown that, on the way to amassing a very large set of characters, starting with exhaustive analyses of limited lexical sets (like the posture verbs system) will be at least as useful as working through a long wordlist in alphabetical order. Covering all the words in a lexical set lessens the likelihood of giving too much weight to flukes, and it reveals typological things along the way. It could give a framework (derivational, semantic, diachronic) within which to explain the behaviour of individual words.

In this small study examples have come up of two stumbling blocks for phylogeny: independent parallel innovation and polymorphy (two equally basic words in one sense, as in Czech dynamic 'sit'). There are no examples of anything resembling a third stumbling block: back-mutation (reversal of a change, e.g. a subgroup protolanguage innovates a word, then one daughter language loses it, thereby appearing never to have had it). How to handle polymorphy is a pressing problem in computational phylogenetics, and I can offer only to have laid out (in the Appendix) a clear example for others to use in modelling.

Overall the danger of parallel development is probably not great in these posture verbs, which differ in their aspectual and derivational morphology from the rest of the lexicon and are not particularly susceptible to lexicon-wide changes, and which are high-frequency items that should be relatively resistant to change. The displacement of dynamic forms by inchoatives is a change that has recurred (it occurs in southern West Slavic and in Macedonian and Bulgarian) but has not been a lexicon-wide one in these languages, so its occurrence in any given lexeme is largely independent of its occurrence in any other lexeme. A change that *is* part of a recurrent process is the derivation of dynamic imperfectives by detransitivisation (reflexivisation), replacing the original iterative intransitives, but since it is aspectually restricted in the East Slavic languages that show it but not in West Slavic, and since in West Slavic (Czech) but not East Slavic it includes derivation from intransitives, it is difficult to say that the same process has occurred. A more delicate subclassification of reflexivisation types may be in order.

The knottiest problem to my mind is the fact that lexical replacements and morphological innovations in individual words are not total and absolute but should really be thought of as shifts in the relative text frequencies of contenders for basic status in a gloss cell. Words commonly have near-synonyms, derivations have alternatives, and additional words and combinations of morphological forms found in one dialect are often known at least passively to speakers of others and of the standard variety — and what happens when a near-synonym, morphological variant, or dialect form becomes standard in some meaning is that it has now become more frequent in that meaning. An example is the perfective dynamic cells in the posture system: both the (unsuffixed) dynamic form and the (suffixed) inchoative could have been present in most Slavic varieties in LPS and later times, and the appearance of one or the other as 'the' form in the dynamic cell could have been a matter of drift in originally subtle differences in text frequency. All of this means that our usual understanding of presence vs. absence of a form from a variety or from a wordlist slot needs to be replaced with a probabilistic measure. Text frequencies are not obviously either transmitted or diffused in the way that words, morphemes, constructions, etc. are, and this means that a statistically sound way of handling polymorphy is part of a larger important problem of describing change. Close philological work over the entire recorded history of each Slavic language should help to clarify the chronology, availability of near-synonyms, and constraints on distribution that will help identify true homoplasies

involving all three of form, function, and constraints and help distinguish them from less diagnostic resemblances.

The database underlying the Appendix will be maintained, updated, and corrected as needed (always with a link and reference to this festschrift), and similar ones provided for other lexical domains, on my website and/or elsewhere, under the heading *Phylogenetica slavica,* for the foreseeable future.

References

Andersen, Henning. 1969. Lenition in Common Slavic. *Language* 45.554–574.

—— 1996. *Reconstructing prehistorical dialects: initial vowels in Slavic and Baltic*. Berlin: Mouton de Gruyter.

Bermel, Neil. 1997. *Context and the lexicon in the development of Russian aspect*. Berkeley-Los Angeles: University of California Press.

Dickey, Stephen M. 2000. *Parameters of Slavic aspect: a cognitive approach*. Stanford, CA: CSLI.

—— 2003. 'Semelfactive' *-noN-* and the Western Aspect Gestalt. *Journal of Slavic Linguistics* 10.29–52.

Forsyth, James. 1970. *A grammar of aspect: usage and meaning in the Russian verb*. Cambridge: Cambridge University Press.

Frček, Jean. 1933. *Euchologium sinaiticum: texte slave avec sources grecques et traduction française*. Paris: Firmin-Didot.

Johnson, Keith. 2008. *Quantiative methods in linguistics.* Malden-Oxford: Blackwell.

Kortlandt, Frederik. 1988. Remarks on Winter's law. In A.A. Barentsen, B.M. Groen and R. Sprenger, eds *Dutch contributions to the Tenth International Congress of Slavists,* 387–396. Amsterdam: Rodopi.

Kuryłowicz, Jerzy. 1947. La nature des procès dits analogiques. *Acta Linguistica* 5.17–34.

—— 1964. *The inflectional categories of Indo-European*. Heidelberg: Carl Winter.

Nakhleh, Luay, Don Ringe and Tandy Warnow. 2005. Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Language* 81.382–420.

Nichols, Johanna. 1993. The linguistic geography of the Slavic expansion. In Robert Maguire and Alan Timberlake, eds *American contributions to the Eleventh International Congress of Slavists,* 377–391. Columbus: Slavica.

—— 2006. Quasi-cognates and lexical type shifts: rigorous distance measures for long-range comparison. In James Clackson, Peter Forster and Colin Renfrew, eds *Phylogenetic methods and the prehistory of languages,* 57–65. Cambridge: McDonald Institute for Archaeological Research.

Nichols, Johanna and Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5).760–820.

Ringe, Don, Tandy Warnow and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100.59–129.

Ross, Malcolm D. 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 Mark Durie and Malcolm 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 event. In Roger Blench and Matthew Spriggs, eds *Archaeology and language I: theoretical and methodological orientations,* 209–261. London-New York: Routledge.

Sadnik, Linda and Rudolf Aitzetmüller. 1955. *Handwörterbuch zu den altkirchenslavischen Texten*. ‘s-Gravenhage: Mouton.

Sławski, Franciszek. 1952. *Słownik etymologiczny języka polskiego*. Krakow: Nakl. Tow. Miłośników Języka Polskiego.

Timberlake, Alan. 1983a. Compensatory lengthening in Slavic, 2: phonetic reconstruction. In Michael S. Flier, ed. *American Contributions to the Ninth Internatioanl Congress of Slavists, 1: Linguistics*, 293–319. Columbus: Slavica.

—— 1983b. Compensatory lengthening in Slavic, 1: Conditions and dialect geography. In V. Markov and Dean S. Worth, eds *From LA to Kiev: Papers on the Occasion of the Ninth International Congress of Slavists, Kiev, September 1983,* Columbus: Slavica.

—— 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.

Townsend, Charles E. 1975. *Russian word formation*. Columbus: Slavica.

Trubachev, O.N. 1967. Iz slavjano-iranskix leksicheskix otnoshenij. *Etimologija 1965,* 3–81. Moscow: Nauka.

Trubachev, O.N., ed. 1974. *Ètimologicheskij slovar’ slavjanskix jazykov.* Moscow: Nauka.

Vaillant, André. 1966. *Grammaire comparée des langues slaves,* vol. 3: *Le Verbe*. Paris: Klincksieck.

Winter, Werner. 1978. The distribution of short and long vowels in stems of the type Lith. *ęsti : vesti : mesti* and OCS *jasti : vesti : mesti* in Baltic and Slavic languages. In Jacek Fisiak, ed. *Recent Developments in Historical Phonology,* 431–446. The Hague: Mouton.

Appendix

The tables here show the triad of verb forms for each posture, and the eight character state encodings. Data columns are headed A, B, etc.; character state columns are headed aa, bb, etc. The citation form is the infinitive except in Bulgarian and Macedonian, where it is a present tense form (these languages have no infinitive). For the past and present forms see text. A blank in one of these columns means that the form is regular and not given in dictionaries. The form chosen as representative of the past stem in the Past column is the *-l* past or participle if it exists and could be found; otherwise the infinitive; otherwise the aorist. Present forms are first singular unless otherwise indicated; 1s - first singular, 3s - third singular, 3p - third plural. Shaded rows or parts of rows give the second of two polymorphic forms for that sense. Optional reflexive clitics (chiefly in Czech and Slovak) are counted as reflexive; the optional prefix in LCS ‘stand’ is counted as no prefix (since this is the earlier and reconstructable state).

Table 7a: Data set for 'sit'

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Static											
CS	sědě-ti	sědě-l-	sěd-i-	*sVd-	1	e:	1	e:	1	e:	1
Russian	sidet'	sidel	sidit	*sVd-	1	i	2	i	2	*e:	1
Belarusian	sjadzec'	sjadzew	sjadzic'	*sVd-	1	e:	1	e:	1	*e:	1
Ukrainian	sydity	sydiv	syžu	*sVd-	1	i	2	i	2	*e:	1
Polish	siedzieć	siedział	siedzi	*sVd-	1	e:	1	e:	1	*e:	1
U. Sorbian	sedźeć		sedźi	*sVd-	1	e:	1	e:	1	*e:	1
Czech	sedět	seděli pl	sedím 1s, sedí 3s	*sVd-	1	e:	1	e: #	1	*e:	1
Slovak	sediet'	sedel	sedí	*sVd-	1	e:	1	e: #	1	*e:	1
Slovene	sedéti	sedèl, sedéla	sedím	*sVd-	1	e:	1	e: #	1	*e:	1
BCS	sjedjeti	sjedje-	sjedi:m	*sVd-	1	e:	1	e:	1	*e:	1
Bulgarian	sedjá	sedjal sedeli; sedjax	sedja; sedi 3s	*sVd-	1	e:	1	e: #	1	*e:	1
Macedonian	sedi	sedevme 1pl, sedeshe 3s	sedi 3s, sedam 1s	*sVd-	1	e:	1	e: #	1	*e:	1
Dynamic perfective											
CS	sěd-ti	sěd-l-	seNd-	*sVd-	1	e:	1	eN	1	none	1
Russian	sest'	sel	sjadu	*sVd-	1	e:	1	eN	1	none	1
Belarusian	sesci	sew, sela	sjádu, sjádze	*sVd-	1	e:	1	eN	1	none	1
Ukrainian	sisty	siv	sjadu	*sVd-	1	e:	1	eN	1	none	1
Polish	siąść	siadł siedli	siądę	*sVd-	1	eN	2	eN	1	none	1
U. Sorbian	sy(d)nyć so	sydźech so	sydnje so	*sVd-	1	e:	1	e:	2	*n	2
Czech	sednout si +	sedl si	sednu 1s	*sVd-	1	e:	1	e:	2	none	1
	posadit se		posadím 1s	*sVd-	1	a	3	a	3	i	3
Slovak	sadnut' (si)	sadol (si)	sadne (si)	*sVd-	1	a	3	a	3	none	1
Slovene	sésti	sel	sédem	*sVd-	1	e	1	eN #	1	none	1
BCS	sjesti	sjeo, sjela	sjedne:m	*sVd-	1	e:	1	e #	1	none	1
Bulgarian	sedna	sédna, sednax; sednal	sedna; sednete	*sVd-	1	e #	1	e #	1	n	2
Macedonian	sedne	sednal	sedne	*sVd-	1	e #	1	e #	1	n	2

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Dynamic imperfective											
LPS	sěd-a-ti	*sěd-a-l-	*sěd-aj-	*sVd-	1	e:	1	e:	1	a	1
Russian	sadit'sja	sadilsja	sazhus'	*sVd-	1	a	2	a	2	i	2
Belarusian	sadzicca		sadzhúsja, sádzicca	*sVd-	1	a	2	a	2	i	2
Ukrainian	sidaty	sidav	sidaju	*sVd-	1	e:	1	e:	1	a	1
Polish	siadać	siadał	siadają	*sVd-	1	e:	1	e:	1	a	1
U. Sorbian	sydać so		syda	*sVd-	1	e:	1	e:	1	a	1
Czech	sedat (si)	sedal (si)	sedám si 1s, sedá 3s	*sVd-	1	e: #	1	e: #	1	a	1
	posazovat se		posazuji se 1s	*sVd-	1	a	2	a	2	ova	3
Slovak	sadat' (si) (dial.	sedat' (si))	sadajú	*sVd-	1	a	2	a	2	a	1
Slovene	sédati	sédal, -a	sedam, 3p -ajo	*sVd-	1	e: #	1	e: #	1	a	1
BCS	sjedati		sjeda:m	*sVd-	1	e:	1	e:	1	a	1
Bulgarian	sjadam	sjadax	sjádam sjada 3s	*sVd-	1	e:	1	e:	1	a	1
Macedonian	sednuva	[sednuvav]	sednuva	*sVd-	1	e:	1	e: #	1	n-uv-a-	4

Table 7b: Data set for 'sit'

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Static											
CS	sědě-ti	sědě-l-	sěd-i-	i	1	impf	1	none	1	no	1
Russian	sidet'	sidel	sidit	i	1	impf	1	none	1	no	1
Belarusian	sjadzec'	sjadzew	sjadzic'	i	1	impf	1	none	1	no	1
Ukrainian	sydity	sydiv	syžu	i	1	impf	1	none	1	no	1
Polish	siedzieć	siedział	siedzi	i	1	impf	1	none	1	no	1
U. Sorbian	sedźeć		sedźi	i	1	impf	1	none	1	no	1
Czech	sedět	seděli pl.	sedím 1s, sedí 3s	i	1	impf	1	none	1	no	1
Slovak	sediet'	sedel	sedí	i	1	impf	1	none	1	no	1
Slovene	sedéti	sedèl, sedéla	sedím	i	1	impf	1	none	1	no	1
BCS	sjedjeti	sjedje-	sjedi:m	i	1	impf	1	none	1	no	1
Bulgarian	sedjá	sedjal sedeli; sedjax	sedja; sedi 3s	i	1	impf	1	none	1	no	1
Macedonian	sedi	sedevme 1pl, sedeshe 3s	sedi 3s, sedam 1s	i	1	impf	1	none	1	no	1
Dynamic perfective											
CS	sěd-ti	sěd-l-	seNd-	none	1	pf	1	none	1	no	1
Russian	sest'	sel	sjadu	none	1	pf	1	none	1	no	1
Belarusian	sesci	sew, sela	sjádu, sjádze	none	1	pf	1	none	1	no	1
Ukrainian	sisty	siv	sjadu	none	1	pf	1	none	1	no	1
Polish	siąść	siadł siedli	siądę	none	1	pf	1	none	1	no	1
U. Sorbian	sy(d)nyć so	sydźech so	sydnje so	n	2	pf	1	none	1	yes	2
Czech	sednout si	sedl si	sednu 1s	n	2	pf	1	none	1	yes	2
	posadit se		posadím 1s	i	3	pf	1	po-	2	yes	2
Slovak	sadnut' (si)	sadol (si)	sadne (si)	n	2	pf	1	none	1	yes	2
Slovene	sésti	sel	sédem	none	1	pf	1	none	1	no	1
BCS	sjesti	sjeo, sjela	sjedne:m	n	2	pf	1	none	1	no	1
Bulgarian	sedna	sédna, sednax; sednal	sedna; sednete	n	2	pf	1	none	1	no	1
Macedonian	sedne	sednal	sedne	n	2	pf	1	none	1	no	1

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Dynamic imperfective											
LPS	sěd-a-ti	*sěd-a-l-	*sěd-aj-	aj	1	impf	1	none	1	no	1
Russian	sadit'sja	sadilsja	sazhus'	i	2	impf	1	none	1	yes	2
Belarusian	sadzicca		sadzhúsja, sádzicca	i	2	impf	1		1	yes	2
Ukrainian	sidaty	sidav	sidaju	aj	1	impf	1	none	1	no	1
Polish	siadać	siadał	siadają	aj	1	impf	1	none	1	no	1
U. Sorbian	sydać so		syda	*aj	1	impf	1	none	1	yes	2
Czech	sedat (si)	sedal (si)	sedám si 1s, sedá 3s	*aj	1	impf	1	none	1	yes	2
	posazovat se		posazuji se 1s	*ouj	3	impf	1	po	2	yes	2
Slovak	sadat' (si) (dial. sedat' (si))		sadajú	*aj	1		1	none	1	yes	2
Slovene	sédati	sédal, -a	sedam, 3p -ajo	aj	1	impf	1	none	1	no	1
BCS	sjedati		sjeda:m	aj	1	impf	1	none	1	no	1
Bulgarian	sjadam	sjadax	sjádam sjada 3s	aj	1	impf	1	none	1	no	1
Macedonian	sednuva	sednuvav	sednuva	n-uv-a-	4	impf	1	none	1	no	1

Table 8a: Data set for 'stand'

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Static											
LPS	stojě-ti	stojě-	stoji-	stoj	1	o	1	o	1	e:	1
Russian	stojat'	stoja-	stoit	stoj	1	o	1	o	1	e:	1
Belarusian	stajac'	staja-	staic'	*stoj-	1	o	1	*o	1	e:	1
Ukrainian	stojaty	stoja-	stoish	stoj-	1	o	1	o	1	e:	1
Polish	stać	sta-	stoj-	*stoj-	1	*o	1	o	1	e:	1
Czech	stát	stá-	stojím	*stoj-	1	*o	1	o	1	e:	1
Slovak	stat'	stál	stojí	*stoj-	1	*o	1	o	1	e:	1
U. Sorbian	stać	sta-	stoji/steji	*stoj-	1	*o	1	o/e	2	?	1
Slovene	stati	stal	stojim	*stoj-	1	*o	1	o	1	?	1
BCS	stajati	staja, stajao	stojim	staj-	1	a	2	o	1	e:	1
Bulgarian	stojà	stojàx	stoìsh 2s, stoja 1s	stoj-	1	o	1	o	1	e:	1
Macedonian	stoi	stoel; stoeja 3pl	stoi	stoj-	1	o	1	o	1	e: #	1
Dynamic perfective											
LPS	*sta-ti	sta-	sta-n-	-sta:-	1	a	1	a	1	none	1
Russian	vstat'	vstal	vstanu	-sta:-	1	a	1	a	1	none	1
Belarusian	ustac'	usta-	ustane	-sta:-	1	a	1	a	1	none	1
Ukrainian	ustaty	usta-	ustanu	-sta:-	1	a	1	a	1	none	1
Polish	wstać	wsta-	wstanę	-sta:-	1	a	1	a	1	none	1
"	powstać	powsta-	[powstanę]	-sta:-		a		a		none	
Czech	vstát	vstá-	vstanu	-sta:-	1	a #	1	a	1	none	1
Slovak	vstat'	vstali	vstane [3s]	-sta:-	1	a	1	a	1	none	1
U. Sorbian	stanyć	stany-	stanu	-sta:-	1	a	1	a ?	1	none	2
Slovene	vstati	vsta-	vstanem	-sta:-	1	a	1	a	1	none	1
BCS	ustati	usta-	ustanem	-sta:-	1	a	1	a	1	none	1
"	stati	sta-	stanem	-sta:-	1	a	1	a	1	none	1
Bulgarian	stàna	stànàx	stana 1s, stànesh 2s	-sta:-	1	a	1	a	1	n	2
Macedonian	stane	stana	stane	-sta:-	1	a	1	a	1	n	2

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Dynamic imperfective											
LPS	*(vъz)staja-ti ?		staj-	-sta:-	1	a	1	a	1	j	1
Russian	vstavat'	vstaval	vstaju	-sta:-	1	a	1	a	1	v	2
Belarusian	ustavac'	ustava-	ustaju	-sta:-	1	a	1	a	1	v	2
Ukrainian	ustavaty	ustava-	ustajesh	-sta:-	1	a	1	a	1	v	2
Polish	wstawać	wstawa-	wstaję	-sta:-	1	a	1	a	1	v	2
"	powstawać	powstawa-	powstaję	-sta:-		a	1	a	1	v	2
Czech	vstávat	vstáva-	vstávám	-sta:-	1	a	1	a	1	v	2
Slovak	vstávat'	vstáva-	vstávaju 3p	-sta:-	1	a	1	a	1	v	2
U. Sorbian	stawać	stawa-	stawa	-sta:-	1	a	1	a	1	v	2
Slovene	vstajati	vstaja-	vstajam, vstavaji	-sta:-	1	a	1	a	1	j	1
BCS	ustajati	ustaja-	ustajem	-sta:-	1	a	1	a	1	j	1
"	stajati	staja-	stajem	-sta:-			1			j	
Bulgarian	stàvam	stavax	stavam	-sta:-	1	a	1	a	1	a	2
Macedonian	stanuva	-uva	-uva	-sta:-	1	a	1	a	1	n-uv-a	3

Table 8b: Data set for 'stand'

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Static						(basic)					
LPS	stojě-ti	stojě-	stoji-	i	1	impf	1	none	1	no	1
Russian	stojat'	stoja-	stoit	i	1	impf	1	none	1	no	1
Belarusian	stajac'	staja-	staic'	i	1	impf	1	none	1	no	1
Ukrainian	stojaty	stoja-	stoish	i	1	impf	1	none	1	no	1
Polish	stać	sta-	stoj-	i	1	impf	1	none	1	no	1
Czech	stát	stá-	stojím	i	1	impf	1	none	1	no	1
Slovak	stat'	stál	stojí	i	1	impf	1	none	1	no	1
U. Sorbian	stać	sta-	stoji/steji	i	1	impf	1	none	1	no	1
Slovene	stati	stal	stojim	i	1	impf	1	none	1	no	1
BCS	stajati	staja, stajao	stojim	i	1	impf	1	none	1	no	1
Bulgarian	stojà	stojàx	stoìsh 2s, stoja 1s	i	1	impf	1	none	1	no	1
Macedonian	stoi	stoel; stoeja 3pl	stoi	i	1	impf	1	none	1	no	1
Dynamic perfective											
LPS	*sta-ti	sta-	sta-n-	n	1	pf	1	none	1	no	1
Russian	vstat'	vstal	vstanu	n	1	pf	1	v-	2	no	1
Belarusian	ustac'	usta-	ustane	n	1	pf	1	v-	2	no	1
Ukrainian	ustaty	usta-	ustanu	n	1	pf	1	v-	2	no	1
Polish	wstać	wsta-	wstanę	n	1	pf	1	v-	2	no	1
"	powstać	powsta-	[powstanę]	n		pf		po-v-		no	
Czech	vstát	vstá-	vstanu	n	1	pf	1	v-	2	no	1
Slovak	vstat'	vstali	vstane [3s]	n	1	pf	1	v-	2	no	1
U. Sorbian	stanyć	stany-	stanu	n	1	pf	1	none	1	no	1
Slovene	vstati	vsta-	vstanem	n	1	pf	1	v-	2	no	1
BCS	ustati	usta-	ustanem	n	1	pf	1	v-	2	no	1
"	stati	sta-	stanem	n	1	pf	1	none	1	no	1
Bulgarian	stàna	stànàx	stana 1s, stànesh 2s	n	1	pf	1	none	1	no	1
Macedonian	stane	stana	stane	n	1	pf	1	none	1	no	1

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Dynamic imperfective											
LPS	*(vъz)staja-ti ?		staj-	j	1	impf	1	none ?	1	no	1
Russian	vstavat'	vstaval	vstaju	j	1	impf	1	v-	2	no	1
Belarusian	ustavac'	ustava-	ustaju	j	1	impf	1	v-	2	no	1
Ukrainian	ustavaty	ustava-	ustajesh	j	1	impf	1	v-	2	no	1
Polish	wstawać	wstawa-	wstaję	j	1	impf	1	v-	2	no	1
"	powstawać	powstawa-	powstaję	j	1	impf	1	po- v-	3	no	1
Czech	vstávat	vstáva-	vstávám	j	1	impf	1	v-	2	no	1
Slovak	vstávat'	vstáva-	vstávaju 3p	j	1	impf	1	v-	2	no	1
U. Sorbian	stawać	stawa-	stawa	j	1	impf	1	none	1	no	1
Slovene	vstajati	vstaja-	vstajam, vstavaji	j/v	2	impf	1	v-	2	no	1
BCS	ustajati	ustaja-	ustajem	j	1	impf	1	v-	2	no	1
"	stajati	staja-	stajem	j		impf		none		no	
Bulgarian	stàvam	stavax	stavam	v (-aj-)	3	impf	1	none	1	no	1
Macedonian	stanuva	-uva	-uva	n-uv-a	4	impf	1	none	1	no	1

Table 9a: Data set for 'lie'

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Static											
LPS	leg-ě-ti	lezhě-	lezh-i-	leg-	1	e	1	e	1	ě	1
Russian	lezhat'	lezhal	lezhit	leg-	1	e	1	e	1	ě	1
Belarusian	ljazhac'	ljazha-	ljazhyc'	*leg-	1	*e	1	*e	1	ě	1
Ukrainian	lezhaty	lezhav	lezhysh 2s	leg-	1	e	1	e	1	ě	1
Polish	leżeć	leżał	leży 3s	leg-	1	e	1	e	1	ě	1
Czech	ležeti	ležel	leží	leg-	1	e	1	e	1	ě	1
Slovak	ležat'	ležalo	leží 3s; ležia 3p	leg-	1	e	1	e	1	ě	1
U. Sorbian	ležeć	ležał	leži	leg-	1	e	1	e	1	ě	1
Slovene	ležáti	ležal	ležím	leg-	1	*e	1	e	1	ě	1
BCS	lèžati	ležala	lèžim	leg-	1	e	1	e	1	ě	1
Bulgarian	leža	lezhax	leži 3s	leg-	1	e	1	e	1	ě	1
Macedonian	lezhi	lezhav 1s	lezhi 3s lezhat 3p	leg-	1	e	1	e	1	ě	1
Dynamic perfective											
LPS	*leg-ti	*leg-l-	*lęg-	*leg-	1	e	1	ę	1	none	1
Russian	lech'	legla	ljag-	leg-	1	e	1	*ę	1	none	1
Belarusian	léhchy	ljahla	ljahu	leg-	1	e	1	*ę	1	none	1
Ukrainian	ljahty	lih, ljahlá	ljazhu	*lęg-	1	*ę	2	*ę	1	none	1
Polish	położyć się	położył się	połozy się	*log-	1	o	3	o	2	none	1
Czech	lehnout si	lehl si	lehne	*leg-	1	*e	1	*e	3	n	2
Slovak	l'ahnut' si	l'ahol si	l'ahne si 3s; l'ahnú si 3p	*leg-	1	*ę	2	*ę	1	none	1
U. Sorbian	lehnyć so	leh(ny)ł so	lehnu so 1s, lehnje 3s	*leg-	1	*e ?	3	*e	3	n	2
Slovene	léči	legel	léžem	*leg-	1	e	1	*e	1	none	1
BCS	leći	lègla	legnem/ležem	*leg-	1	e	1	*e	1	none	1
			ležem								
Bulgarian	legna	legnax	legna	*leg-	1	e	1	*ę ?	1	n	2
Macedonian	legne	legnav 1s; legnale Pl	legne 3s	*leg-	1	e	1	*ę ?	1	n	2

Language	Citation form	Past	Present	A Root	aa	B Past root V	bb	C Pres root V	cc	D Past suffix	dd
Dynamic imperfective											
LPS	*lěgati	*lěga-	*lěgaj-/*lěž- [< *lěg-j-	*lVg-	1	ě	1	* ě	1	a	1
Russian	lozhit'sja	*log-i-	*log-i-	*lVg-	1	o	2	o	2	i	2
Belarusian	lazhycca	*log-i-	lazhusja	*lVg-	1	o	2	o	2	i	2
Ukrainian	ljahaty	ljahav	ljahaju	*lVg-	1	*ę	3	*ę	6	a	1
Polish	kłaść się	*klad-	*klad-	*klad-	2	a	4	a	3	none	3
Czech	lehati si	lehalo	lehají 3p	*lVg	1	* ě #	5	* ě #	4	a	1
Slovak	líhat' si	líhal si	líhajúci ppl	*lVg	1	* ě #	5	* ě #	4	a	1
U. Sorbian	lěhać so		lěham [so]	*lVg	1	* ě	5	* ě	4		1
Slovene	légati	legali	légam	*lVg	1	* ě #	3	* ě #	4	a	1
BCS	lijègati		lijèžem	*lVg	1	* ě	5	* ě	4	a ?	1
Bulgarian	ljagam	ljagax	ljagam	*lVg	1	* ě	5	* ě	4	a	1
Macedonian	legnuva			*lVg	1	e #	3	e #	5	n-uv-a-	4

Table 9b: Data set for 'lie'

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Static											
LPS	leg-ě-ti	lezhě-	lezh-i-	i	1	impf	1	none	1	no	1
Russian	lezhat'	lezhal	lezhit	i	1	impf	1	none	1	no	1
Belarusian	ljazhac'	ljazha-	ljazhyc'	*i	1	impf	1	none	1	no	1
Ukrainian	lezhaty	lezhav	lezhysh 2s	*i	1	impf	1	none	1	no	1
Polish	leżeć	leżał	leży 3s	i	1	impf	1	none	1	no	1
Czech	ležeti	ležel	leží	i	1	impf	1	none	1	no	1
Slovak	ležat'	ležalo	leží 3s; ležia 3p	i	1	impf	1	none	1	no	1
U. Sorbian	ležeć	ležał	leži	i	1	impf	1	none	1	no	1
Slovene	ležáti	ležal	ležím	i	1	impf	1	none	1	no	1
BCS	lèžati	ležala	lèžim	i	1	impf	1	none	1	no	1
Bulgarian	leža	lezhax	leži 3s	i	1	impf	1	none	1	no	1
Macedonian	lezhi	lezhav 1s	lezhi 3s lezhat 3p	i	1	impf	1	none	1	no	1
Dynamic perfective											
LPS	*leg-ti	*leg-l-	*lęg-	none	1	pf	1	none	1	no	1
Russian	lech'	legla	ljag-	none	1	pf	1	none	1	no	1
Belarusian	léhchy	ljahla	ljahu	none	1	pf	1	none	1	no	1
Ukrainian	ljahty	lih, ljahlá	ljazhu	mut.	2	pf	1	none	1	no	1
Polish	położyć się	położył się	połozy się	none	1	pf	1	none	1	yes	2
Czech	lehnout si	lehl si	lehne	n	3	pf	1	none	1	yes	2
Slovak	l'ahnut' si	l'ahol si	l'ahne si 3s; l'ahnú si 3p	n	3	pf	1	none	1	yes	2
U. Sorbian	lehnyć so	leh(ny)ł so	lehnu so 1s, lehnje 3s	n	3	pf	1	none	1	yes	2
Slovene	léči	legel	léžem	none	2	pf	1	none	1	no	1
BCS	leći	lègla	legnem/ležem	n	3	pf	1	none	1	no	1
			ležem	mut	2						
Bulgarian	legna	legnax	legna	n	3	pf	1	none	1	no	1
Macedonian	legne	legnav 1s; legnale pl	legne 3s	n	3	pf	1	none	1	no	1

Language	Citation form	Past	Present	E Pres suffix	ee	F Aspect	ff	G Prefix	gg	H Refl	hh
Dynamic imperfective											
LPS	*lěgati	*lěga-	*lěgaj- / *lěž- [< *lěg-j-	aj	1	impf	1	none	1	no	1
Russian	lozhit'sja	*log-i-	*log-i-	i	2	impf	1	none	1	yes	2
Belarusian	lazhycca	*log-i-	lazhusja	i	2	impf	1	none	1	yes	2
Ukrainian	ljahaty	ljahav	ljahaju	aj	1	impf	1	none	1	no	1
Polish	kłaść się	*klad-	*klad-	none	3	impf	1	none	1	yes	2
Czech	lehati si	lehalo	lehají 3p	aj	1	impf	1	none	1	yes	2
Slovak	líhat' si	líhal si	líhajúci ppl	aj	1	impf	1	none	1	yes	2
U. Sorbian	lěhać so		lěham [so]	aj	1	impf	1	none	1	yes	2
Slovene	légati	legali	légam	*aj ?	1	impf	1	none	1	no	1
BCS	lijègati		lijèžem	none	3	impf	1	none	1	no	1
Bulgarian	ljagam	ljagax	ljagam	*aj ?	1	impf	1	none	1	no	1
Macedonian	legnuva			n-uv-a-	4	impf	1	none	1	no	1

8 *Greenberg's Indo-Pacific hypothesis: an assessment*

ANDREW PAWLEY

1 Introduction

In 1971 Joseph Greenberg published evidence for his Indo-Pacific hypothesis, which proposed that there is a genetic relationship between all the non-Austronesian language families of Melanesia and the Halmahera and Timor regions of the Indonesian archipelago, together with the languages of Tasmania and most if not all the languages of the Andaman Islands.[1] The main evidence consisted of 84 sets of resemblant words ('Indo-Pacific etymologies') plus some resemblances in grammatical elements, almost every set being represented in at least three of the 14 groups which he treated, provisionally, as primary branches of Indo-Pacific (IP).

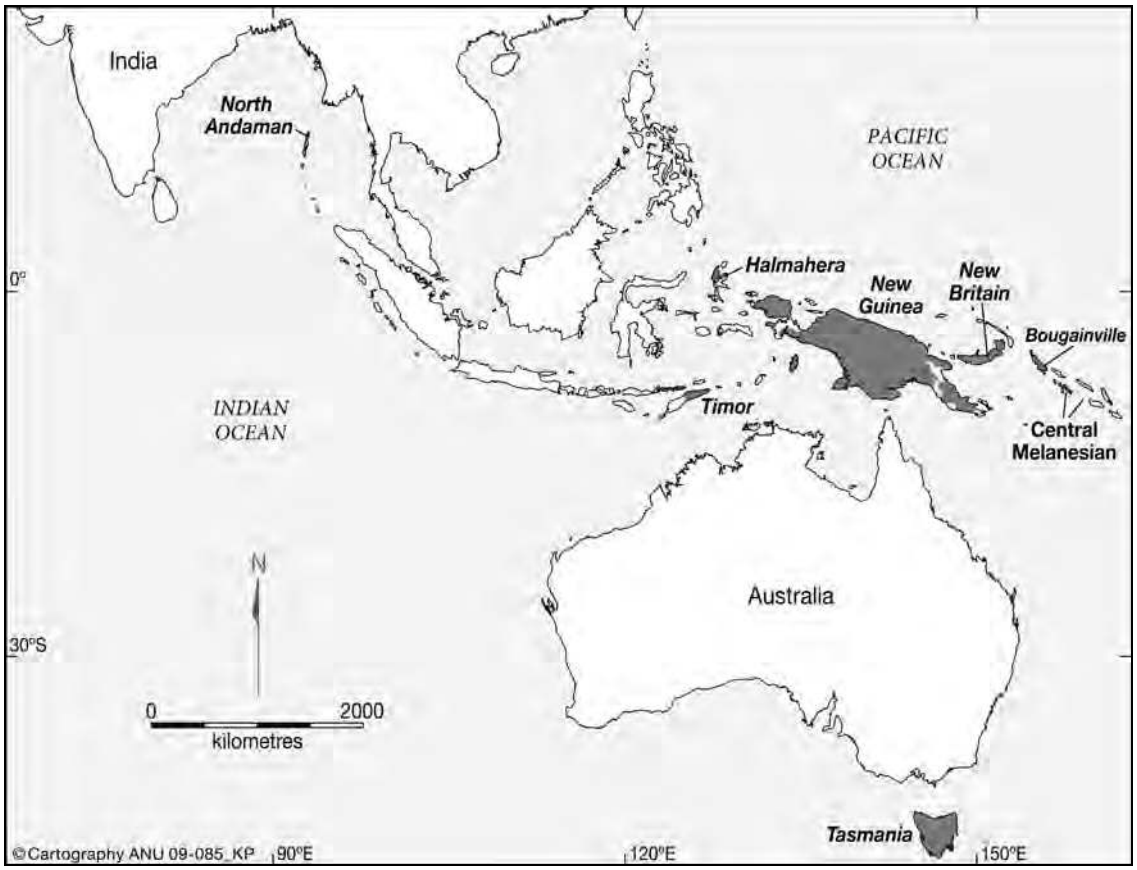
Twelve of these 14 groups are found in a more or less continuous area of the equatorial zone that extends from 123 degrees E to 166 degrees E, from Timor and Halmahera in the west, through the large (2400 km long) island of New Guinea, to New Britain, New Ireland and the Solomon Islands in the east. The dominant language family in this area, except for New Guinea, is Austronesian. The non-Austronesian languages, numbering between 700 and 800 (Wurm 1982; Wurm and Hattori 1981–83), are often collectively termed 'Papuan', traditionally without any implication that this label refers to a genetic grouping. The Andaman Islands lie much further west, south of Burma and the Bay of Bengal, at around 92 degrees E. Tasmania, situated to the south of the Australian mainland, is separated from all the other groups by a vast distance. Recently, Whitehouse et al. (2004) have argued that the Kusunda language of Nepal should be added to Indo-Pacific.

In this paper I will try to do four main things: (1) assess the merits of Greenberg's arguments for the Indo-Pacific hypothesis, (2) point out why specialists have up till now largely shirked this task, (3) evaluate weaker alternatives to the full-scale hypothesis, in which only some of the putative primary subgroups are included, and (4) reflect on the circumstances and chain of reasoning which persuaded Greenberg that he had a fairly good

[1] It gives me great pleasure to contribute to this volume honouring Malcolm Ross. It is my privilege and good fortune over the last 20 years to have had Malcolm as a colleague and friend and as a partner in a number of research projects in Austronesian and Papuan historical linguistics. I remain in awe of his accomplishments. Roger Blench, Beth Evans, Judith Blevins, Edgar Suter and Matthew Spriggs provided valuable comments on a draft of this paper.

case for a hypothesis when contemporary specialists in Papuan historical linguistics find these arguments unconvincing.

Beginning in the mid-1950s, Greenberg spent much time over a dozen years compiling putative Indo-Pacific etymologies, patiently tracking down data from obscure published and unpublished sources and entering materials in a notebook using his 'multilateral' or 'mass comparison' method, to be discussed in §5 below. In addition to the 84 sets of putative Indo-Pacific cognates he put together hundreds of other sets of resemblant words restricted to the putative subgroups of Indo-Pacific. Tentative findings were first reported in two unpublished papers, Greenberg (1958), where the 14 groups were defined and Greenberg (1960), where the full-scale Indo-Pacific hypothesis was first proposed. He continued to add to his materials until 1968, when he submitted the paper that appeared in 1971. According to Croft (2005:xviii), he examined some 350 lexical entries plus grammatical comparisons for about 800 non-Austronesian languages (plus some 50 neighbouring An languages for controls).[2]



Map 1: Location of the putative major subgroups of Indo-Pacific
(See §2 for details of the 14 groups)

[2] While Greenberg's diligence in tracking down data was extraordinary, I doubt if he could have obtained 350 items for as many as 800 non-Austronesian languages in the 1950s and 1960s. There are probably fewer than 800 distinct Papuan languages in all, and for perhaps 200 of these there was almost no documentation during that period. However, Greenberg sometimes had wordlists for multiple dialects of single languages.

Greenberg (1971:854) writes that:

> I believe that the evidence presented here is sufficient to establish the point that [in the Indo-Pacific region] the vast majority of non-Austronesian languages outside of Australia, on which judgment is still reserved, have a common origin…. For Tasmanian the relative paucity of data produces a somewhat weaker case than in other instances. Still what evidence we have does point in this direction.

He adds that 'My hope is that the present study will help to hasten the long overdue demise of the notion of Papuan as merely a scrapheap of assorted languages bound together by the negative characteristic of being non-Austronesian. May the comparative study of this major linguistic stock, which has been so strangely neglected, … finally come into its own.' (1971:854)

The Indo-Pacific hypothesis has had a rather curious fate. Greenberg's arguments for Indo-Pacific have been summarily dismissed by various specialists as unconvincing, but none of the critics has provided a detailed assessment of the evidence. At the same time, the Indo-Pacific grouping is often mentioned in handbooks and encyclopaedias describing language families of the world, and is sometimes presented there as a more or less established stock (e.g. Ruhlen 1991). From time to time linguists doing comparative typological work (e.g. Viberg 1984) cite the hypothesis as if it were reasonably well supported, as do scholars in other disciplines, including population geneticists (e.g. Cavalli-Sforza 2000; Thangaraj et al. 2003) and historians (Manning 2006). This degree of acceptance is certainly not based on a rigorous assessment of the evidence — for no such assessment has been provided so far — but is surely due to the eminence of Greenberg himself, whose groundbreaking work on linguistic universals and on the classification of African languages made him one of the most influential linguists of the 20th century.

In recent decades research in archaeology and population genetics has greatly advanced our knowledge of the history of human settlement of Island SE Asia and the Pacific Islands. At present the most widely supported view among population geneticists and archaeologists is that the first successful colonisation of Asia, beyond the Levant, by *Homo sapiens* did not occur until between 70,000 and 50,000 years ago (Mellars 2006a, b). The Andaman Islanders, genetically, represent a clade of modern humans with no close relatives elsewhere (Thangaraj et al. 2003; Thangaraj et al. 2006). They appear to be a long isolated population deriving from the first modern human colonisation of South and Southeast Asia.[3]

It is now known that by at least 45,000 years ago modern humans were in Borneo, then still part of mainland SE Asia (Barker et al. 2005). By 45 to 40 millennia ago (and possibly several millennia before that) they had crossed Wallacea and reached Sahul, the Australia-New Guinea continent (Groube et al. 1986; O'Connell and Allen 2004; O'Connor 2007). Indeed as early as 40,000 BP people had made the sea crossings to New Britain and New Ireland (Leavesley and Chappell 2004; Pavlides and Gosden 1994; Specht 2005; Torrence et al. 2004) and by about 30,000 BP they were in Bougainville (Specht 2005; Spriggs 1997; Wickler and Spriggs 1988). The spread southwards across Australia, probably initially following the coasts, was quite rapid. By 35,000 BP the remote southwest corner of what is now the island of Tasmania was populated (Cosgrove, Allen and Marshall 1990; Mulvaney and Kamminga 1999). The genetic evidence indicates that, aside from some input from Austronesian speakers from SE Asia within the last three millennia or so, the

[3] At present, archaeological dates for the Andaman Islands go back no more than about 2000 years but comparatively little archaeological work has been done there.

current Papuan-speaking peoples of New Guinea and Island Melanesia derive from these foundation populations (Friedlaender 2007).

Thus, if there was a common ancestor shared by the languages of the Andaman Islands, the Papuan languages of Melanesia, and the languages of Tasmania it is likely to have been more than 40,000 years in the past. If there was a common ancestral language shared by all the Papuan languages of New Guinea, the Bismarcks and the Solomons it was probably more than 30,000 years ago.[4] Genetic and archaeological evidence indicate that there was little if any interaction between the founder populations of the Solomons and the rest of Melanesia in the period between initial settlement of the Solomons archipelago some 30 millennia ago (Friedlaender 2007; Friedlaender et al. 2008; Spriggs 1997) and the arrival of Austronesian speakers.

These chronologies do not rule out the possibility that the Indo-Pacific hypothesis is correct. It may be that the first early modern human expansion into SE Asia and Sahul was carried by very small populations speaking languages of a single family. However, the issue is not whether the Indo-Pacific languages share a remote common ancestor — indeed, all human languages may do so — but whether there is reasonable linguistic proof of common origin. The chronology for the first colonisation of Australia and Melanesia raises questions about what kinds of shared linguistic residues, if any, are likely to have survived after 40 millennia. From what we know of rates of replacement of particular kinds of lexical and grammatical roots, only a few dozen words have half-lives[5] of more than 2000 years and the only elements that have half-lives of more than 20,000 years are some personal pronouns and a handful of lexical items, such as words for certain body-parts and kinship terms, and a few other concepts, probably fewer than 20 in all.[6]

It is unlikely that any cognate sets for items outside of the small hard core would have survived for 40 millennia, and even if they did, phonological changes would very likely have obscured their common origin. In this connection the fate of Greenberg's (1987) Amerind hypothesis (apparently independently developed in the 1950s by Morris Swadesh and Sydney Lamb) is instructive. Greenberg assigned to Amerind all 60 or so established stocks of Native American languages other than Eskimo-Aleut and Athabaskan.

---

4 These early dates were not known to Greenberg when he was formulating the Indo-Pacific hypothesis. Possibly he would not have been influenced by them because he believed in treating comparative linguistic evidence independently of non-linguistic evidence (Croft 2005:xii).

5 To say that the half-life of a word (more exactly, a particular lexical form-meaning pairing) is 2,000 years means that, in any language that has that it, the word has a 50 per cent chance of persisting (with the same meaning) for that period of time.

6 For a number of language families it has been shown that words for certain concepts are extremely persistent while words for other concepts are less stable. For discussion of Indo-European basic vocabulary see, e.g. Kruskal et al. (1971), Pagel (2000), Pagel and Meade (2006), Pagel, Atkinson and Meade (2007). Dyen et al. (1967) determined the cognation rates of words for 196 meanings in 89 Austronesian languages. Only 10 meanings show cognation rates of above 50 percent: *two, four, give, eye, we, louse, father, mother, to die, to eat*. Another 15 meanings show rates between 50 and 30 percent: *one, three, ashes, stone, nose, to hear, to drink, new, thou, ye, fruit, name, ear, liver, tree.* Another 45 show cognation rates of between 29 and 10 percent. Rates for the remaining 126 meanings fall below 10 percent. Austronesian is a family that lends itself well to determining absolute, as well as relative replacement rates in basic vocabulary. The dates at which Proto Austronesian and its major interstages were spoken are reasonably well-established because the Austronesian expansion left a well-marked archaeological trail (Bellwood 1997; Green 2003; Kirch 2000; Pawley 2002). This allows approximate retention rates to be calculated for particular lexical items reconstructed for Proto Austronesian (spoken about 5000 BP), Proto Malayo-Polynesian (about 4000 BP) and Proto Oceanic 3400–3100 BP).

Archaeology and genetics indicate that the Americas were probably not settled until about 15,000 years ago, and if these first settlers brought a single language one would expect the residue recoverable from comparison of the several hundred Amerind languages to be considerably larger than the residue recoverable for Indo-Pacific. Yet it has proved very difficult to persuade specialists that Amerind is a valid genetic stock (Campbell 1988, 1997; Campbell and Poser 2008; Mithun 1999; Nichols and Peterson 1996). It is not that the specialists are being obtuse but rather that their standards of proof are more demanding than Greenberg's were.

To estimate probabilities of chance resemblances ideally one needs data on the frequency of particular sounds in particular positions in all the relevant languages (Ringe 1992, 1996, 1999), data that are not available for Amerind. Instead, one must make do with approximations based on averaging data for certain languages. Following this procedure, Ringe (1996:152) 'finds no evidence whatsoever that the putative cognate sets in Greenberg's 'Amerind Etymological dictionary' represent anything other than chance'.

Although I will conclude that there is no good case for the full-scale Indo-Pacific hypothesis, Greenberg deserves credit for seeing, as early as the 1950s, that many of the small, disparate groups of non-Austronesian languages in the New Guinea area are probably related and, above all, for assembling a body of resemblant items that at least provide a basis for discussion. The material assembled by Greenberg includes some lexical and grammatical resemblances that indicate a common origin of most of the languages in six of the 14 primary groups that he posited. Subsequent work has shown that these languages belong to the language family now generally termed 'Trans New Guinea' (TNG). With some 400–450 member languages TNG is probably the third most numerous family in the world, after Niger-Congo and Austronesian.

However, Greenberg's failure to recognize that the TNG languages collectively amount to a single first-order witness rather than six severely weakens his arguments for the Indo-Pacific hypothesis as a whole. It turns out that by far the best evidence for Indo-Pacific consists of agreements among diverse branches of TNG. By contrast, the case for a genetic relationship between the North Andaman languages and the Tasmanian languages, on the one hand, and any of the Papuan groups of New Guinea, New Britain, Bougainville and the Solomons is extremely weak, the quantity and quality of the resemblances falling well within the range of chance.

Greenberg's material contains a few items that hint at a remote genetic relationship between the TNG family and certain other language families of New Guinea, and a similar connection may exist between certain non-TNG families of New Guinea and certain languages of New Britain. However, these do not amount to anything like a convincing case.[7]

2 Greenberg's subgrouping of Indo-Pacific languages

Greenberg's assumptions about the internal relationships of Indo-Pacific languages were central to his weighting of agreements in lexicon and grammar. He distinguished 14 major subgroups, which were assumed to have equal status as putative primary branches of Indo-Pacific. Most groups are defined by particular group-specific lexical resemblances (i.e. resemblances that go beyond the 84 putative Indo-Pacific etyma) and in some cases by

[7] For more recent discussion of evidence concerning distant relationships among the diverse Papuan families see Dunn et al. (2002), Dunn et al. (2005), Reesink (2005).

particular grammatical features. A few of the groups were based on geography: they subsume different genetic stocks found in the same small region, which as a precautionary measure Greenberg treated as a single unit. Within most of the primary subgroups he distinguished further branches.

For some of the primary and secondary subgroups Greenberg lays out the evidence. In other cases he simply refers to published or unpublished evidence without citing details of resemblances. He offered words of caution:

> This subgrouping is not exhaustive and is in some respects at least quite tentative … Such a degree of uncertainty is only reasonable at this stage' (1971:809).

Greenberg's subgroups, with the names and abbreviations he uses plus brief notes on the extent of the evidence he cites, are listed below. The order follows a directional pattern, moving roughly from west to east in the tropical zone, and then to Tasmania in the south.

The Andaman Islands

1. Andaman (AN). Greenberg observes that the Andaman languages fall into two groups that have not been shown to be related. He includes in Indo-Pacific only the larger, North Andaman group, which occupies almost all of Great Andaman Island and which consists of two closely related dialect clusters. He reserves judgment on whether the two South Andaman languages are related to the North Andaman group.

Indonesian archipelago

2. Timor and Alor (TA). Greenberg had data for only four of the non-Austronesian languages in the Timor-Alor region, at the eastern end of the Lesser Sundas chain, namely Abui, Bunak, Makasai and Oirata. Ninety-two comparisons are given in support of this grouping, of which 16 are also part of the list of 84 Indo-Pacific etymologies. He recognised two branches within the Timor-Alor group.

3. Halmahera (HA). The Papuan languages of north Halmahera 'form an obvious group … so that no demonstration is necessary' (Greenberg 1971:815). They divide into a southern group, made up of Ternate and Tidore, and a northern group consisting of some 10 languages, including Loloda and Tobelo.

New Guinea mainland

4. West New Guinea (WNG). About 40 languages are named in this group group, all located at the western end of New Guinea, in parts of the Bird's Head and the Bomberai Peninsula. They are divided into four subgroups: (1) a large group of almost 20 languages including Etna Bay and Mairasi, (2) a group of about 10 languages including Madi, Tehit and Waken, (3) a group of four languages including Maibrat, and (4) Kapaur, Baha and Kovas. Twenty-seven etymologies were cited linking Etna Bay with Cowan's (1957) West New Guinea family. (More problematically, Cowan also included Halmahera and Timor languages in his West New Guinea family.)

5. Southwest New Guinea (SWNG) or Marind-Ok. Five subgroups are distinguished. Four of these, Tirio, Marind, Ok, and Awyu are located in south-central New Guinea, close to the Papua New Guinea border. The fifth, Kukukuku, located in Gulf and Morobe provinces, is a very tentative inclusion. About 60 supporting lexical comparisons are cited.

6. Southern New Guinea (SNG) or Kiwaic. In this group of some 50 languages Greenberg distinguishes seven branches, all spoken on or near the coast of southern New Guinea, in Western Province and the Torres Straits and on Frederik Hendrik Island. The subgroups are not given names but the largest are those that contain Kiwai and its immediate relatives, and Jei and its immediate relatives. Miriam, of the Eastern Torres Straits, belongs to SNG. The westernmost group is centred on Fredrerik Henrik Island. Seventy lexical comparisons support the putative SNG group.

7. North New Guinea (NNG). Under this rubric Greenberg combines several very divergent groups, scattered around north-central New Guinea on both sides of the West Papua-PNG border and in the Sepik provinces. He identifies these groups as Sentani, Tami, Arapesh, Murik, Monumbo (Manambu) and Ndu-Kwoma, along with some isolates. Fifty lexical comparisons plus pronominal agreements are cited in support of NNG.

8. Central New Guinea (CNG). This is the largest of the 14 groups assigned to Indo-Pacific. Greenberg, influenced by Wurm (1964), who had tentatively posited a large 'Central and Northeast New Guinea Phylum', recognised three primary branches: (i) Kapauku-Baliem (also known as Ekagi-Dani) in the western highlands of West Papua, (ii) a Central Highlands group (similar to Wurm's East New Guinea Highlands Stock) and (iii) Huon (in the region of the Huon Peninsula, just north and east of the Central Highlands group). The Central Highlands group in turn divides into several groups, including those known nowadays as Engan (including Huli, Mendi, Kewa and Ipili), Chimbu-Wahgi, and Kainantu-Gorokan. No supporting cognate sets are cited other than those in the main Indo-Pacific list. For the smaller groups he refers to the published work of others.

9. Northeast New Guinea (NENG) or Madang. To this Greenberg assigns 30 or so languages of Madang Province. He remarks that the nucleus of such a group was recognised by Ray (1919) and that '[t]he unity of this group is quite obvious' (1971:834). No etymologies are cited.

10. East New Guinea (ENG). This putative group has more than 80 members located in the southeastern region of New Guinea. Thirty-seven lexical comparisons and four pronominal agreements are cited in support of ENG. Greenberg recognises ten subgroups, including the larger groups he calls Mailu, Binandere, and Koita, along with Dimpa, Kovio, and Elema.

The Bismarck Archipelago

11. New Britain (NB). Greenberg had data for five languages, all of Central and East New Britain. He recognised that there are two or more very divergent groups on this large (450 km long) and mountainous island, prone to cataclysmic volcanic eruptions, but treats them as a unit for geographic reasons. Four languages, comprising the Baining-Sulka group, are clearly related. He was uncertain whether the fifth, Uasi, had any special relationship to other New Britain languages and he had no data on a language, Idne, said to be spoken in the far west of the island. No etymologies are cited in support of NB.

The Solomon Islands

12. Bougainville (BO). There are two highly divergent groups, one in the north and one in the south. Each contains four languages (some with diverse dialects). Greenberg considers that they have enough in common to justify a Bougainville subgroup (2005:203) but he cites no cognate sets in support of this claim.

13. Central Melanesia (CM). There are four non-AN languages of the central Solomons 'which seem to constitute another subgroup' (Greenberg 1971:816) that he calls Central Melanesian. He also tentatively assigned to this group certain languages of Reefs/Santa Cruz Islands, situated almost 1000 km to the east, mainly on grounds of resemblances in the pronouns. Fifty-two lexical and nine pronominal etymologies are cited for CM.

Tasmania

14. Tasmania (TS). The only data consist of brief and highly problematic notes made by 19th century colonials. Following Schmidt (1952) Greenberg recognises five languages (or dialects): a northern language and four others that appear to be more closely related to each other. No supporting lexical data are cited, other than those in the Indo-Pacific list.

Following Cowan (1957, 1960), Greenberg (1971:839) speculates that Halmahera, Timor–Alor and West New Guinea may constitute a 'supergroup', on the basis of some agreements in grammatical features.

Unclassified languages of New Guinea (UNG). Greenberg also referred to, but left unclassified, a number of very small groups and isolates in New Guinea and to one isolate spoken on New Ireland.

3 The reception of the Indo-Pacific hypothesis. Why have specialists largely ignored it?

Why has the Indo-Pacific hypothesis received little attention from specialists in the relevant language groups? In the four decades since Greenberg's main publication on this subject there have been a handful of brief assessments by specialists, consisting of just a few sentences, and all have rejected the evidence as unconvincing (among these are Laycock 1975a; Pawley 1998, 2005a; Ross 2005). It is noteworthy that in Foley (1986), a book devoted to the Papuan languages, there is no reference to the Indo-Pacific hypothesis and that in another book on the Papuan languages, Wurm (1982:6, 30) simply acknowledges Greenberg's proposal in three sentences.[8] There have been a couple of papers that briefly examine Greenberg's arguments for particular subgroups of Indo-Pacific (Franklin 1973; Voorhoeve 1975). This reception stands in contrast to the lively and extensive debates generated by Greenberg's African and Amerind classifications.

As far as I know, only one slightly more extended review of on the Indo-Pacific hypothesis has appeared: a five page commentary by Wurm (1975a:925–929). However, this commentary contains much hedging and little discussion of nitty-gritty details. The main points could have been made in half a page. They are that:

(i) Greenberg made several claims about relationship between diverse Papuan groups that now, in the light of better data than he had, appear not to be demonstrably related. This in turn casts serious doubt on the value of his evidence for the claim that Tasmanian and Andaman are also related to the Papuan languages.

[8] I suspect that neither Wurm nor Foley wished to offend a respected colleague and chose not to air their disagreements. Foley in particular has close links to Stanford.

(ii) The case for Tasmanian is particularly weak. Among the grammatical items, there is a single pronominal resemblance, in the 2SG forms. Eighteen of the 84 lexical etymologies include Tasmanian items but these are not convincing.

(iii) The number of resemblances exhibited by the Andaman group seems, at first blush, to be significantly higher. There are three pronoun items that show a resemblance to pronouns found in certain other groups, plus the past tense marker *k*. But Wurm observes that the pronominal agreements are really much weaker than the foregoing statement implies, because they are divided among disparate groups: the 1SG agreement is with 'West Papuan', the 2SG agreement is with 'East Papuan', and the 1PL agreement is with yet another set of languages. Thirty of the 84 Indo-Pacific etymologies are represented in Andaman, and the resemblances are chiefly with W. Papuan and Timor-Alor languages. Wurm suggests that there may be an ancient substratum in the Papuan area that is linked to the Andaman languages.

There are, I believe, several reasons why scholars have been reluctant to attempt a detailed assessment of Greenberg's Indo-Pacific proposal. First, a thorough review would be very time-consuming. The Indo-Pacific hypothesis is in fact a cluster of many hypotheses about genetic groupings, each of which has more or less independent status and would have to be assessed one by one. Apart from the full-scale Indo-Pacific grouping there are the 14 proposed subgroups, many of which are highly problematic.

A second reason has already been alluded to above: Greenberg did not separate the wheat from the chaff. He had the makings of a good case for linking several Papuan stocks in New Guinea with each other and with certain languages of the Timor-Alor area, but did not separate this from the much flimsier case for including the languages of Tasmania, the Andamans, Halmahera and Island Melanesia. Unsurprisingly, some readers faced with these more far-reaching and weakly supported claims, were inclined to ignore the rest.

Third, the pool of specialists who are more or less competent to review the evidence has always been very small. The total number of linguists actively working on the historical study of any or all of the Papuan families has probably never exceeded ten or twelve at any one time (the peak was between about 1965 and 1975) and since the early 1980s has been considerably fewer. Among these scholars, only one or two have had Papuan historical linguistics as their primary research field. As for the Andaman and Tasmanian languages, the situation is worse.

Fourth, the timing of Greenberg (1971) was unlucky. His thunder was largely stolen by discoveries in Papuan studies that were being reported in the 1960s and 1970s. The idea that some of the diverse, small Papuan groups might be related was in the air during the 1950s, as can be seen in the writings of Capell (1948–49), Cowan (1957), Loukotka (1957) and Wurm (1954). However, the kinds of arguments put forward in these works were chiefly typological, and we can now see that in many cases the early tentative proposals did not stand up.

Beginning in the late 1950s, Stephen Wurm, at The Australian National University, initiated a long term program of field surveys and comparative research on the Papuan languages of New Guinea and Island Melanesia. He was soon joined by several collaborators — both departmental colleagues and PhD students. In the early 1960s Wurm published a series of papers giving typological and lexicostatistical evidence for a family of around 50 languages located in the central highlands of Papua New Guinea, which he called the East New Guinea Highlands Stock (later called a Phylum) (Wurm 1960, 1964,

1965). In the mid-1960s several proposals concerning long-range connections among diverse groups in New Guinea were put forward, such as Wurm's (1965, 1971) Central New Guinea Macro-Phylum.[9] These were based mainly on typological agreements, the lexicostatistical agreements between widely separated groups being too low (2–5%) to be significant.

Then at the end of the decade McElhanon and Voorhoeve (1970) cited about 90 sets of resemblant lexical items shared by several widely separated groups: namely a group located in central and south-central New Guinea (Voorhoeve 1968), another in the Huon Peninsula area of central north New Guinea (McElhanon 1967, 1970) and a third, the small Binandere family, situated in the southeast of the island. They coined the name 'Trans New Guinea phylum' for this widespread group. McElhnanon and Voorhoeve used a method very similar to Greenberg's multilateral method (§5.1) to assemble possible cognates, without attempting to work out regular sound correspondences. Among their sets of resemblant forms were several pronouns and other items of core basic vocabulary.

This first, restricted version of the TNG hypothesis was soon dramatically extended — indeed McElhanon and Voorhoeve (1970) had suggested that their TNG Phylum would turn out to be related to the groups in Wurm's Central New Guinea Macro-Phylum. Within a few years, the central highlands family and various other groups, including the Timor area languages, had been added to TNG, so that almost 500 languages, or about 70 percent of all non-Austronesian languages of the region, were assigned to this family in its most extended form (Wurm ed. 1975; Wurm 1982; Wurm et al. 1975).

As it happens, proponents of the extended TNG hypothesis in the 1970s did not make good use of the evidence they had. The case they made for TNG was poorly made and far from convincing because it relied too much on typological resemblances, and provided no systematic phonological and lexical reconstruction. All informed reviewers were highly sceptical (Foley 1986; Haiman 1979; Heeschen 1978; Lang 1976). However, more recent work has yielded more solid evidence for TNG (with a membership not quite as extensive as that proposed in Wurm (ed. 1975)). Accounts of the history of the TNG hypothesis are given in Pawley (1998, 2005a) and Ross (2005).

In the 1970s several other major genetic groupings besides TNG were posited by the ANU group. These included the Sepik-Ramu Phylum, to which were assigned almost 100 languages of north central New Guinea (Laycock and Z'graggen 1975), the Torricelli Phylum, consisting of some 47 languages of the Torricelli Ranges and nearby regions of the Sepik and Ramu Provinces of Papua New Guinea (Laycock 1975b) and the East Papuan Phylum, said to subsume all 20 or so non-Austronesian languages of Island Melanesia (Wurm 1975b). Recent opinion is that neither the Sepik-Ramu nor the East Papuan groupings stand up, the evidence for Sepik-Ramu being flimsy (Foley 2005) and that for East Papuan even more so (Ross 2001). However, these speculative groupings were included in the influential *Atlas of languages of the Pacific* (Wurm and Hattori 1981–83) and outsiders have often assumed that they are well-supported genetic groups.

---

[9] The extended TNG hypothesis had in fact been roughly foreshadowed in a 1965 report by the Voegelins, where they proposed to unite the East New Guinea Highlands Stock with a Huon Peninsula group, the Binandere group, the Ok group of central New Guinea, and the Dani group of the SW New Guinea Highlands. They also threw in the Ndu family of the Sepik, which is not TNG. However, the Voegelins relied on the data and proposals of Greenberg, Wurm and other sources and did not analyse the evidence further.

Finally (and this is arguably the most important single reason for the lack of a detailed assessment) there are major methodological difficulties in evaluating the evidence. Greenberg's method of multilateral comparison yielded a rather small body of impressionistic resemblances between form-meaning units. In such cases, unlike claims about cognation that rest on regular sound correspondences, one cannot appeal to rigorous and reliable criteria to assess a claim of common origin. The claims crave statistical testing for significance but such testing would be time-consuming and few linguists are well equipped to do it.

Such methodological concerns recur in most of the brief assessments of Greenberg (1971), for example, that of Laycock (1975a:57):[10]

> To date it can safely be said that there is no real evidence to link the [non-Austronesian] languages of New Guinea with any other linguistic groups … In particular Greenberg's Indo-Pacific hypothesis … is not only far from proven, but also based on inadequate and insufficiently analysed data (for example, comparisons are too frequently made of items within larger groups of languages — such as the Trans New Guinea Phylum that are already known to be related, so that there is little support for the wider relationships postulated.)

4 Contemporary views of the genetic classification of Papuan languages

Contemporary views of the classification of the non-Austronesian languages of Melanesia and Southeast Asia, and the languages of Tasmania, differ in various ways from Greenberg's. A detailed account of contemporary views would require a separate paper. A thorough review of the Tasmanian data was undertaken by Crowley and Dixon (1981). Research on the Andaman languages is reviewed by Blevins (2007, to appear). Wurm (1975a, 1982) surveyed work on the 'Papuan' languages up to the mid 1970s. The most comprehensive recent classification of the 'Papuan' languages, based mainly on pronominal paradigms, is in an unpublished paper of some 200 pages by Ross (2000), whose main findings are summarised in Ross (2005) and in Pawley (2005b, 2007). Among other studies that treat some of the groups accepted or proposed by Greenberg are the following. Z'graggen, in a number of works (e.g. Z'graggen 1975), confirmed and extended the Northeast New Guinea (Madang) group. Ross (2001) examined Wurm's (1982) hypothesis that the various non-Austronesian families of Island Melanesia (Melanesia excluding New Guinea) belong to a diverse East Papuan phylum. Foley (2005) argued against the Sepik-Ramu hypothesis (the core of Greenberg's North New Guinea group). Pawley (1998, 2005a, b) and Ross (1995, 2005) argue in support of the Trans New Guinea hypothesis. Voorhoeve (2005) discusses inheritance and diffusion among certain of the groups making up Greenberg's Southwest New Guinea group and Reesink (2005) does the same for West New Guinea. Dunn et al. (2002, 2005) examine the distribution of a large selection of typological characteristics across the various families that Wurm (1982) had assigned to 'East Papuan', in an attempt to find traces of ancient common origin or diffusion.

---

[10] However, the words 'already known to be related' in this quote are unfair to Greenberg. The 'larger groups' that Laycock refers to, such as the Trans New Guinea Phylum and Sepik-Ramu Phylum, were not proposed in print until the early 1970s and even then were not well supported (see below). Greenberg (1971) was submitted in 1968, some three years before it was published as part of a large multi-authored volume. I am confident of this, first, because none of the chapters in that volume contain references dated later than 1968 and because some other contributors to this volume told me they had a deadline of 1967 or 1968.

The most important differences with Greenberg's views concerning genetic relationships to emerge are listed below:

(i) It is now clear that (leaving aside certain problematic single languages) almost all the languages in Greenberg's groups 5 (Southwest New Guinea or Marind-Ok), 8 (Central New Guinea), 9 (Northeast New Guinea or Madang), 10 (East New Guinea), and part of group 6 (Southern New Guinea or Kiwaic) belong to a single large family, Trans New Guinea (Pawley 1998, 2001, 2005a, b; Ross 1995, 2000, 2005). Of particular importance is the fact that the non-Austronesian languages of group 2, Timor and Alor (along with those of Pantar), geographically isolated from the rest, have fairly strong claims to be assigned to TNG. Indeed, Ross (2000), on somewhat slender pronominal evidence, specifically assigns them to a subgroup that has other members on the New Guinea mainland, around the Bomberai Peninsula.

(ii) Although all the languages Greenberg assigned to group 5, Southwest New Guinea (or Marind-Ok) belong to TNG, this set of languages is not now regarded as forming a subgroup (Pawley 2005a; Voorhoeve 2005).

(iii) Group 6, Southern New Guinea, is not regarded as a genetic group but is divided into several families, one of which is TNG. Ross (2000) tentatively includes Kiwai and its immediate relatives in TNG but not the rest of group 6.

(iv) Group 7, Northern New Guinea, is not regarded as a genetic group but is divided into several families and a few isolates (Foley 2005; Ross 2000). A few of the languages Greenberg assigned to NNG are TNG.

(v) Although all the languages Greenberg assigned to group 8, Central New Guinea, are now assigned to TNG, they are not viewed as otherwise forming a subgroup. Indeed, the large Central Highlands branch of CNG posited by Greenberg (following Wurm) is not now regarded as forming a subgroup of TNG. On the contrary, the Central Highlands languages fall into several groups that on present evidence appear to be first-order branches of TNG.

(vi) Although all the languages Greenberg assigned to group 10, East New Guinea, belong to TNG, they are not now viewed as forming a subgroup.

(vi) Group 11, comprising the New Britain languages, divides into at least two families (Ross 2000, 2001), a possibility that Greenberg acknowledged.

(vii) Group 12, Bougainville. Ross (2000, 2001) finds no case, on the pronominal evidence, for uniting the two highly divergent groups, one in the north and one in the south.

(viii) Group 13, Central Melanesian. Ross (2000, 2001) finds very weak evidence for relating the four non-AN languages of the central Solomons. Ross and Næss (2007) have shown that the Reef Islands language is not 'Papuan'. It belongs to the Oceanic subgroup of Austronesian and, by association this holds for the Santa Cruz languages, which are its immediate relatives.

(ix) Group 14. Crowley and Dixon (1981) conclude that there were at least six distinct languages represented in the meagre data recorded from Tasmania, but probably between eight and twelve. The materials consists of 200 to 300 words for some South-east lects and much smaller amounts for other lects. The only clear grammatical data available are forms for 'I' and 'you' in a few languages.

Although some of the languages are clearly related, the data do not permit the conclusion that all the Tasmanian languages are related. And,

> [a]lthough Tasmanian languages seem typologically similar to languages of the Australian family [in their phonologies], there are insufficient cognates [read 'resemblant forms'] to justify an even tentative hypothesis of genetic relationship (Crowley and Dixon 1981:395).

5 On the lexical evidence for Indo-Pacific

5.1 Greenberg's etymologies

In various places in his writings Greenberg makes the point that the first step in the comparative method is working out which languages to compare, i.e. which languages are likely to be genetically related. He regarded his method of 'multilateral comparison' as an efficient way of carrying out this first step. He described the method as one that 'looks at everything at once' (2005:94). Word lists are arranged so that one's eye scans a few words across many languages, rather than many words across a few languages. That is, there is simultaneous comparison of languages and lexical items from the full range of languages and language families under consideration. Greenberg makes the following observation about the value of the method as a discovery procedure.

> Most important of all, perhaps, is that where more than one family is represented, … the contrast between the relatively numerous and qualitatively superior resemblances between related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances. (Greenberg 2005:42)

This observation is surely true but there is a certain irony in it, when we consider the quality of the evidence for Indo-Pacific.

As an example of the power of the multilateral method Greenberg lists words for diverse European languages, organised so that all the Germanic languages are contiguous, likewise the Celtic languages, the Romance languages, and so on, and writes

> In Table 7 I have listed a few basic words for twenty-five languages of Europe. The number of ways of classifying twenty-five languages, even without specifying subgroupings, is 4639×10^{19}, that is, over a quintillion. Yet the correct classification and even subgroupings and intermediate groupings (e.g. Balto-Slavic) are apparent from just a cursory glance at two or three words (2005:94).

(One can accept Greenberg's main point here but it should be noted that he has organised the table to make this easy. A random listing would take more than a cursory glance to sort out.)

That is all very well for Indo-European and its major branches but it is clear multilateral comparison does not work so well when the groups are, at best, only very distantly related — otherwise, of course, there would not be such a level of disagreement among scholars as we find. Multilateral comparison relies on there being enough resemblant items shared by a pair (or larger set) of languages to decisively indicate common origin without the time-consuming work of establishing regular sound correspondences. For the putative high-order subgroups of Indo-Pacific we cannot compile tables comparable to those available for Indo-European because the number of resemblant forms in basic vocabulary is much, much smaller. And this is the critical difference. The problems are to know (a) what counts

as resemblant items, (b) how many such items are enough and (c) how to distinguish cognates from chance similarities and borrowings.

The pitfalls of trying assessing resemblances without knowing the phonological history of the languages are illustrated by #38 'head'. Greenberg compares CM languages that have forms of the type of Savo *mbatu* with Bunak (Timor) *ubul* and Yela Dne (Rossel Island, SE Papua New Guinea) *mbara*. But apart from the phonological differences, the CM forms are Austronesian loans: reflexes of Proto Oceanic *b^{w}atu 'head' are widely reflected (as *mbatu*, etc.) in the Oceanic languages of the Solomon Islands.

Greenberg is wont to quote statistics indicating that the chances of certain sets of resemblances occurring by chance are infinitesimally remote. One must take these estimates with a large grain of salt, because all too often there are counterexamples. What are the chances that English and Maori, two unrelated languages, would show marked similarities in the numerals 2, 3, 4? They do. Compare English *two* (Scots *twa*), *three*, *four*, with Maori *rua*, *toru*, *whaa* (where *wh* is a bilabial fricative). All the Germanic languages show comparable likenesses to almost all the Polynesian languages.

As a sample of the difficulties posed by the proposed Indo-Pacific etymologies consider comparisons #56 to #59. In #56, for the meaning 'old', resemblant forms are cited from witnesses in four far-flung groups: Andaman (four languages) *tam* and *taum*, Halmahera (one language) *timono*, Central New Guinea (two languages) *tamana, tamon*, and the Solomon Islands (one language) *tam*. In #57, headed 'to plait', resemblant forms are cited from just two groups: Andaman (Biada *tepi*) and Halmahera (Tobelo *tapi*). In #58, for 'to push', forms are cited from two groups: Andaman (Bogijieb *tera*) and Halmahera (Tobelo *tila*). In #59, headed 'rain', forms are cited from four groups: Tasmania (four languages have *moka* 'water'), WNG (two languages have *moka* 'wet', NNG (seven languages have a range of forms such as *mayk*, *mac*), and SWNG (two languages have *mauka* 'water).

In the absence of any knowledge of the historical phonology of any of the languages cited, what can be said about these resemblant items? We can note the formal similarities and ask what is known about the stability of terms meaning 'old', 'to plait', 'to push', etc. We can speculate on how likely it is for a few languages out of 750 to retain such resemblant forms after 40,000 years of separate development, and try to calculate how likely it is that such resemblances could have developed independently ('by chance') in different groups. But without a rigorous statistical analysis using fair and reasonable criteria there is simply no way of separating the wheat from the chaff other than one's personal judgment.

I consider that, among Greenberg's 84 Indo-Pacific etymologies, about 23 contain a core of convincing resemblant items. A list of the most promising etymologies is given below. The sets are numbered as in Greenberg's list but I have greatly abbreviated the material. Instead of citing long lists of forms from those of Greenberg's subgroups that we now assign to TNG, I cite a reconstruction attributable to an early stage of TNG (here labelled simply 'pTNG'). In reconstructed forms C = consonant, V = indeterminate vowel. Most of the reconstructions are drawn from Pawley (2005a, n.d). Particular resemblant forms are cited from Andaman and Tasmania languages but for other non-TNG groups I merely note, in most cases, that a particular subgroup is represented in the set of resemblances. Putative resemblances cited by Greenberg that seem very far-fetched are discarded from the comparisons listed below.

Table 1: The most promising of Greenberg's Indo-Pacific etymologies

above (1)	PTNG *op(V)
arm (4)	PTNG *mbena, AN *ben* 'shoulder-blade'
bark (7)	PTNG *ka(nd,t)apu, AN *kait*, *kaic*, TS *kite*, NNG
bone (12)	pTNG *kondaC, TS *teni*
come (18)	PTNG *ma(n)-
die (21)	PTNG *kumV-
ear (23)	PTNG *damV, NB, NNG
earth (24)	PTNG *ma(l,n)a
eat (25)	pTNG *na-, BO, NNG
egg (26)	PTNG *maŋgV 'round, compact object', AN *molo, mula, mule*
female (28)	PTNG *pan(V), BO
fire (30)	PTNG *inda, AN *at*, TS *to*, *toi*
hair (33)	PTNG *iti, AN *de*, HA ??
husband/male (42)	PTNG *ambi
lip/mouth (45)	PTNG *ambe, AN *pe*, *pa*
louse (47)	PTNG *niman, NNG, NB, UG
moon (51)	PTNG *kal(a,i)m
nose (58)	PTNG *mundu
older sibling (63)	PTNG *nan(a,i)
stay (65)	PTNG *mVna-
star (71)	PTNG *bay, TS p*oe*, ENG, CNG, UNG
stone (73)	PTNG *kambu(CV), BO, NNG
tongue (76)	PTNG *me[l,n]e, TS *mena*, BO, UNG

All or almost all of the remaining Indo-Pacific etymologies, close to three-quarters, can be discarded as 'chaff'. A good many of the putative cognate sets represent meaning-form pairings that typically have quite short half lives, e.g. 'arrow', 'beautiful', 'bush, forest', 'buttocks', 'to dance', 'fog', 'mud', 'to plait', 'thing' 'to push', 'to walk', 'white', 'yellow'. The fact that Greenberg is able to find roughly similar forms for these concepts in diverse Indo-Pacific groups, that have independent histories for the past 30 millennia, must weaken our confidence in the reliability of the method. A good many resemblances are only included by allowing the semantic net to be cast very wide. For example, the set of 'earth' includes forms glossed 'bottom', 'underneath', 'mud', 'land'; under 'walk', are included forms meaning 'leg, foot'; under 'ear' are included verbs 'to hear'; and so on.

My view is that none of the lexical resemblances between North Andaman and Tasmanian languages and between members of either of these groups and other groups assigned to Indo-Pacific are due to common origin. There are three reasons for this conclusion:

1. The Andaman and Tasmanian populations have been isolated from each other and from the Papuan speaking peoples of Melanesia for at least 40,000 years (see discussion in §1). Everything known about rates of lexical replacement in large language families indicates that the shared lexical residues left after 40 millennia are likely to be very, very meagre and entirely confined to a small core of basic vocabulary, probably fewer than 20 words. Furthermore, phonological changes would very likely have obscured the common origin of almost all the surviving cognates.

2. There is no compelling collection of resemblances in the hard core basic vocabulary. Only one or two noteworthy agreements are found in that domain: Tasmanian *mena* 'tongue', TNG *me[l,n]e, and perhaps AN *pe* 'lip, TNG *ambe 'mouth'. These isolated likenesses are not enough to make a case.
3. Given the very large number of languages compared the overall number of lexical resemblances is small and not above chance levels. Among these likenesses are some that are too good to be true — very similar forms for meanings that are not core basic vocabulary. There are superficially impressive resemblances between many language families that are not generally regarded as related. For instance, enthusiastic amateurs (and occasionally professionals) have come up with hundreds of look-alikes shared by Semitic and Austronesian, by Japanese and Austronesian, by Quechua and Austronesian, and even by Niger-Congo and Austronesian.

What about Greenberg's lexical evidence for relating what we now know to be the TNG family to other putative Indo-Pacific subgroups from Bougainville and the Central Solomons? Essentially the same objections apply to this evidence as to the case for relating Andaman and Tasmanian languages to the languages traditionally known as Papuan. As noted earlier, the archaeological record suggests that, following initial settlement of what was then the island of Greater Bougainville some 30,000 years ago there was little or no contact between New Guinea populations and populations in Bougainville and the Solomons until the advent of Austronesian speakers around three millennia ago and the genetic record is consistent with this conclusion.

The lexical evidence for connecting TNG with certain other languages of the New Guinea mainland is, I think, slightly stronger. For example, forms resembling the very stable TNG etyma *niman 'louse' and *na- 'to eat' occur in a number of non-TNG languages of New Guinea. But there is no space here for a detailed assessment of this evidence.

6 On the grammatical evidence for Indo-Pacific

Greenberg (1971:842ff.) cites agreements in 11 grammatical features (and alludes to others). He regarded these, especially certain pronominal agreements, as the strongest part of his evidence for including the various non-New Guinea groups in Indo-Pacific. The trouble is, again, that the strongest agreements are between members of TNG. As recent work has confirmed, one can reconstruct for pTNG a complete paradigm of independent personal pronouns and part of a set of verbal suffixes marking subject person-and-number and some other fragments of morphology. The problem is to make a case for reconstructing grammatical features to a stage earlier than pTNG.

The following table of TNG independent pronouns is based on Ross (2005:29), as slightly modified in Pawley (2005a:89):

Table 2: Proto TNG independent pronouns[a]

	1	2	3
singular	na	ŋga[b]	[y]a, ua
plural (i-grade)	ni	ŋgi[b]	?
plural (u-grade)	nu		
dual (i-grade)	ni(l,t)i	ŋgi(l,t)i[b]	i(l,t)i
dual (u-grade)	nu(l,t)i		
non-singular	nja		

[a] Ross also reconstructs an inclusive suffix *-m- 'plural' and *-p- 'dual' (2005:29).
[b] A case can be made for reconstructing the initial consonant as *k rather than *ŋg.

The following is a critical summary of Greenberg's account of the grammatical evidence for Indo-Pacific.

1. First person singular pronouns. He notes that two sets of forms are widespread.

(i) *n*-forms 'absolute (independent)'. The reconstruction of pTNG *na '1SG independent' is generally accepted. This accounts for the occurrence of *n*- forms in TA, CNG, SWNG, SNG, NENG, and ENG. (Within TNG, *na reflexes are absent from the Madang and SE Papuan groups.)

Outside of TNG *n*-forms are found in:

West New Guinea: The Konda-Jahadu and Kapaur groups have *n*- (Kampong Baru *neri* '1SG', *eri* '2SG', Tarof *ne(iga)* '1SG' *va(iga)* '2SG').

North New Guinea: Ndu has *n*-forms (Maprik *unə*, Kwoma, Mayo *an*).

Bougainville: Telei *na*, Nasioi and Koromiva *n*- 'my'.

Central Melanesia: Savo *-ni* '1SG obj.', *n*- + '1 object marker'. (The 'cognates' in Santa Cruz languages, which are now classified as Austronesian, are invalid.)

(ii) *t*- forms for subject and object. Within TNG, these are found in TA (Makasai *ani* '1SG absolute', *asi* possessive), Kainantu: Benabena *nani* absolute, *-te* possessive (and other Tairora group languages). In NENG *t*-forms are widespread for both absolute and possessive uses.

Outside of TNG, *t*- forms are found in WNG. About half of WNG languages have *t*-forms for subject and object, and the other half have *n*- forms.

2. Second person singular pronouns. Greenberg finds that 'over a large part of New Guinea' (1971:844) there is an opposition between first person *n* (usually *na*) and second person *k*- (usually *ka*). (Here he has recognised the TNG pattern. pTNG *ka or *ŋga '2SG' is well attested.) Where the *na/ka* pattern does not predominate the most common second person pronoun is *ngi* or *ni*. 'I suspect that *ngi* is original and has frequently become *ni* either by direct phonetic change or under the influence of first person singular *n*' (1971:844).

Beyond TNG, 2SG forms with initial *n* or *ng* occur in:

Andaman: Biada *ngol*, Onge *ngii.*

Halmahera: Galela *no* '2SG subject', *ni* '2SG object, *ngona* '2SG independent'.

West New Guinea: Amberbaken, Madik, Karon *nan*, etc.

North New Guinea: Tanggum *nu*, Murusapa *na*, Anaberg *nə*.

New Britain: Baining *ngi*, Taulil *nggi*, *ngginggi*, Uasi *nini*.

Central Melanesia: Savo *no*, Bilua *ngo*, Baniata *no*.

Tasmania: All dialects *ni(na).*

3. First person plural pronouns. Greenberg found *ni* is widespread in groups that we now assign to TNG. pTNG *ni and *nu '1PL' are well attested.

Similar forms are found in certain other languages of New Guinea:

West New Guinea: *ni(ti)* '1pl.excl.' in Solowat, Higo and congeners.

North New Guinea: Anaberg *ni*, Tanggium *nai*, Sko, Sagke *ne*, Ndu *nanə*, etc.

Unclassified New Guinea: Rossel (Yela Dnyi) *nu-* 'our'.

Beyond New Guinea, *n-* initial forms are found in North Bougainville (Telei, Nasioi *nii*, *ni* 'our') and Halmahera (*na* marks 1pl object inclusive in most HA languages).

At first blush these resemblances in first and second person pronouns between TNG and members of other groups listed in (1–3) above look impressive. However, there are a number of grounds for caution. Ross (2005:50) is critical of Greenberg's application of the multilateral method to pronouns in his Amerind work, treating pronouns as individual forms rather than as part of paradigmatic sets. In his Indo-Pacific study Greenberg cites paradigms where possible but does not hesitate to include resemblant pronominal forms that come from different paradigmatic sets.

There is a general problem in evaluating formal resemblances among pronouns. Rhodes (1997) argues that functional pressures restrict the range of phonological features used to mark pronominal contrasts, thereby increasing the likelihood of chance similarities among pronoun forms. One such factor is that pronouns or pronominal affixes are typically backgrounded in discourse. This means, among other things, that they tend to be short (singular markers almost always a single syllable) and unstressed. Three problems must be solved for backgrounded items to be communicatively effective:

(a) identification: one must be able to tell when one is hearing a morpheme of the relevant type, e.g. a pronoun, not a noun.

(b) differentiation. One must be able to distinguish among members of this class.

(c) ease of pronunciation. One must be able to pronounce the items with relative lack of attention.

These factors stand in partial conflict and produce a range of optimal pronominal systems.[11] The ease of pronunciation consideration strongly favours use of unmarked segments, i.e. the more common or most common segments in pairs or larger sets of phonemes. Rhodes cites work by Gordon (1995) who using a sample of 62 languages of diverse families found that consonants and vowels occur in pronominal systems with the following frequencies (percentages rounded out).

[11] Rhodes' account of the differentiation and identification problems refers to quite complex factors that allow a variety of optimal systems and I will say almost nothing about these here. The differentiation problem favours systems that maximize acoustic distinctness but not in a way that reflects any sound symbolic link between one of the persons and one of the classes of sounds.

Table 3: Frequencies of consonants and vowels in pronominal systems across 62 languages

consonant	% of languages	vowel	% of languages
n	93	a	98
m	75	i	90
k	71	u	69
t	68	o	56
y	53	e	52
w	43		
h	40		
ŋ	39		
s	37		
r	37		
…			
ñ	19		

That is to say, this factor favours the use of small inventories of segments in pronominal systems. Among consonants, *n, m, k* and *t* are highly favoured. Among vowels, *a* and *i* are highly favoured.

Nichols and Peterson (1996) use a larger sample. In their study the 1SG pronoun has *n* as the initial C in 37/173 languages (or 20.8%), 2SG has *n* as the initial in 23/173 languages (13.3%).

Given that singular pronouns are generally monosyllabic and that *n and *k are highly favoured consonants in pronoun systems it would seem we need to treat Greenberg's pronominal evidence for Indo-Pacific with some caution. Indeed there is other evidence showing that the chances of two languages independently developing 1SG, and 2SG pronouns beginning with the same consonant are by no means miniscule. It happens that Trans New Guinea 1SG, 2SG and 3SG independent pronouns have close matches in some languages of the Afro-Asiatic, Algonquian and Austronesian families. The following table compares the well attested pTNG forms with the independent pronouns of Hausa (Afro-Asiatic) and SW Ojibwe (Algonquian) and with preverbal subject pronouns in two Austronesian languages of Vanuatu: Mera Lava and Raga.

Table 4: Singular pronouns in languages of four unrelated families

	pTNG	Hausa	Ojibwe	Mera Lava	Raga
1SG	*na	ni	ni:n	na, no	na
2SG	*ga	kai	ki:n	ko	go
3SG	*ya	shi	wi:n	a	k-ea

The Hausa 1SG and 2SG forms continue Proto Chadic forms that are similar. The Proto Algonquian singular pronouns are: 1sg *ni:la, 2sg *ki:la, 3sg *wi:la, with P-Alg *l > n in Ojibwe, merging with reflexes of P-Alg *n (J. Blevins pers. comm.).

The Mera Lava and Raga subject pronouns continue Proto Austronesian (PAn) independent forms with a number of changes. In the first singular forms the initial *n* is not original. PAn *aku '1SG' became *au in Proto Oceanic (POc), with irregular loss of *-k. In

a number of Vanuatu and SE Solomons languages POc *au became *nau, possibly because the *n of a preceding transitive marker was reanalysed as part of the object pronoun (the same as the independent forms). The independent singular pronouns were then adopted as preverbal subject markers, with some phonological reduction ensuing. *nau reduced to *na* and *no* in the two languages in question, thus coming to closely resemble the pTNG form. The PAn 2SG root was a disyllable, *kaSu, which normally took a prefix *i that marked independent pronouns. The Mera Lava and Raga forms continue *kaSu regularly, with *S lost and *au becoming o, thus creating a monosyllabic form that closely resembled the pTNG form. In the third singular the PAn form *(si-)ia (yielding POc *ia) is quite similar to pTNG *ya, and this is continued in Mera Lava and Raga with some irregular developments. There are other cases of Austronesian languages in Indonesia that have independently developed three singular pronouns closely resembling those attributed to pTNG.

While these observations do not rule out the possibility that some or all of the pronominal resemblances between TNG and non-TNG groups are due to common origin they show that there is a reasonable chance that some or all of the resemblances, specially those between geographically well-separated groups, may also be due to non-genetic factors.

4. Timor-Alor and Halmahera agree in having a first person inclusive plural pronoun in *p. The Timor-Alor witness is a TNG language but there is no good reason to think that this is even an old TNG feature, so this resemblance is unlikely to be a shared inheritance.

5. Third person plural. *d* and *t* forms occur in:

   New Britain: Taulil, Butam, Sulka *ta*.

   Bougainville: Siwai *at*, Galeli *idu*. NNG: Sko *tea*, Sangke *te*, *ndu*, (*n*)*di*.

   Weak. Such resemblances between three widely separated groups are likely to be due to chance.

6. Suffixes on verb marking subject person and number. This is characteristic of most TNG groups. Greenberg notes that the structural type also occurs in Nimboran, of NNG, but such a structural resemblance in a language close to TNG languages is of little value. Any formal agreements between particular suffixes are confined to TNG.

7. Pronouns (i) prefixed to noun to indicate possessor, (ii) prefixed to verb to indicate object. These features are characteristic of the TNG family only. Given SOV order, it is not surprising that object pronouns precede the verb.

8. In three New Guinea groups (SWNG, SNG, CNG), all now assigned to the TNG family, certain tenses are marked by subject-tense portmanteau suffixes in which:

   (i) second and third persons are identical in non-singular dual and plural

   (ii) first person differs from non-first person by a vowel change which is the same for plural (and for dual if there is one). Recurrent variants are *a/i*, *e/i* and *i/e*. These two features, and especially *a/i* variation, may well be old TNG features but they are not attested in Indo-Pacific groups other than TNG.

9. A plural marker on nouns, *mana* or *mVnV*, occurs in some languages in three of Greenberg's groups: Timor-Alor (in Abui), Central New Guinea (Moni), and East New Guinea (in Binanderean). These are all TNG groups.

10. Marking of grammatical gender (or noun classes). Grammatical gender, done by vowel alternations, is a feature of a number of Indo-Pacific groups. In most groups the masculine vowel is more front than the feminine. Greenberg considers this correlation to be a major piece of evidence for his hypothesis. He discusses at some length gender marking in Marind, a TNG language of south central New Guinea. Marind has four genders: 1. masculine human, 2. feminine human and animals, 3. inanimate, 4. inanimate. The most basic pattern is: *e* masculine singular, *u* feminine singular, *a* inanimate class 1, *i* inanimate class 2 + plural of masculine and feminine This pattern is manifested in some nouns such as *anem* 'man, *anum* 'woman', *anim* 'people' but more widely in adjectival agreement with nouns.

Within TNG gender-marking is virtually confined to the south central New Guinea area and there are no strong grounds for attributing it to pTNG. However, gender marking is found in several other Papuan groups. Halmahera uses consonant variation for this purpose. Gender marking is widespread in NNG languages, e.g., Monambo of the Sepik region, has a five gender system with feminine singular *u*, neuter singular *i*, and three consonantal markers. Taulil and Butam of New Britain have masculine *a*, feminine *e*, neuter *i* (a striking resemblance to Marind and Monambo) and plural *ta.*

In Bougainville Nasioi has contrasts like *nuring* 'son', *norang* 'daughter', *naung* 'husband', *naang* 'wife', where *i* marks masculine and *a* feminine.

The specific correlations that Greenberg points to are indeed striking and may be the shadowy remnants of an ancient shared history. However, without a cross-linguistic survey of the kind that has been done for pronoun forms it is hard to evaluate the chances of these sorts of resemblances arising independently. My impression is that vowel alternations are quite widely used to mark gender contrasts in determiners, pronouns and nouns.

11. Past tense marked by a suffix containing a velar consonant. This feature is found in some members of at least four TNG groups (SWNG, SNG, CNG, NENG) and such a distribution yields a fairly promising case for reconstructing pTNG *-k '(remote) past'. We also find *-ka* in most North Andaman languages and some Halmahera languages and forms containing *k* or *g* in some NNG languages and in Bilua, a Central Melanesian language. However, once again, in the absence of other, more convincing evidence for connecting AN, HA, NNG and CM it is difficult to place much weight on this resemblance. Given that suffixes tend to erode and that velar stops are not the most stable of consonants, the chances of any language retaining a past tense suffix based on a velar consonant for 30 or 40 millennia would seem to be very small.

To sum up, the grammatical evidence includes several morphological agreements that support a TNG group, namely items 1–3, 6–8, and 11, and perhaps 9. There is some shadowy evidence for connecting TNG with certain other New Guinea area groups and isolates. It would hardly be surprising if TNG shares a common ancestor with some other languages of the New Guinea mainland at a time depth of between 10 and 15 millennia, recent enough for a few traces to remain.

The evidence for relating either Tasmanian and Andaman to any of the other groups is negligible. The few resemblances are best viewed as accidental. The same assessment applies to resemblances between Bougainville and Central Melanesian and any of the other groups.

7 Why was Greenberg persuaded?

Finally, we are left with this question: Why was Greenberg, an extremely erudite and astute scholar, and the author of some highly regarded papers on the methodology of historical linguistics, persuaded that he had a pretty good case for Indo-Pacific when his critics are unimpressed by the evidence? Several factors can be readily discerned.

First, Greenberg seriously underestimated the chances of different languages independently developing resemblant pronoun forms and, more generally, resemblant lexical forms. Greenberg drew a very long bow in assigning putative cognates to his 'Indo-Pacific etymologies', allowing great latitude both in respect of formal and semantic variation. This weakness would no doubt have been corrected had he followed application of the multilateral method with statistical tests for significance. His reluctance to take this further step remains a puzzle.

Second, there is the subgrouping issue. Greenberg deserves credit for recognising a number of grammatical and lexical agreements that support what we now call the Trans New Guinea family. However, the evidence he compiled for uniting groups 2, 5, 8–10 and part of 6, i.e. the TNG languages, is much stronger than the evidence for any wider grouping among the 14 putative subgroups of Indo-Pacific. To the extent that there are resemblances among groups 1, 3, 4, 6, 11–14 and between these and the TNG groups, they are few in total and flimsy in quality. If Greenberg had counted the numbers of resemblances across his subgroups the differences would surely have been obvious but he did not provide any statistical arguments and it is possible that that he did not see the patterning. At any rate, with the benefit of hindsight we can see that his failure to identify the Trans New Guinea languages as a single primary unit in his subgrouping hypothesis, rather than as representing several coordinate subgroups, led him to overvalue the importance of agreements between the TNG groups as evidence for a wider Indo-Pacific stock.

Third, he did not try to support his etymologies by seeking recurrent sound correspondences, either within or between particular subgroups. Given the scope of Indo-Pacific, we can hardly blame Greenberg for not investigating sound correspondences — for most of the putative subgroups he could not have made much progress in such a task with the fragmentary data at his disposal and even with excellent data the job of analysing correspondences for all of the groups would be beyond any single person. However, it is not especially difficult to demonstrate recurrent sound correspondences between the better-known TNG languages. Greenberg himself could have done so for the languages which figure most prominently in his etymologies, had he chosen to undertake this step in the comparative method

Greenberg was critical of the categorisation of historical linguists into 'lumpers' versus 'splitters', arguing that the number of groups related under a hypothesis should not be an issue. But surely the central issue has always been the quality of the evidence. The difference is that lumpers are satisfied with a lesser standard of proof than splitters. It seems that, in the case of Indo-Pacific, Greenberg forgot his own wise advice, cited earlier in this paper, and which I repeat here:

> ... where more than one family is represented, … the contrast between the relatively numerous and qualitatively superior resemblances between related languages, compared to the sporadic and qualitatively poorer resemblances among unrelated languages, becomes readily apparent. In this way the presence of unrelated languages provides a control for distinguishing mere chance from genetically significant resemblances. (Greenberg 2005:42)

I remain astonished that such a hugely experienced and perceptive scholar did not take a more cautious and critical view of the evidence before him. My hunch is that Greenberg's early successes in relating African groups made him addicted to the search for long range relationships and led him to take a less critical view of the evidence than he should have. Great scholars are not immune to hubris.

References

Barker, Graeme, Tim Reynolds and David Gilbertson, eds. 2005. *The human use of caves in Peninsular and Island Southeast Asia.* Special Issue. *Asian Perspectives* 44(1).

Bellwood, Peter. 1997. *The prehistory of the Indo-Malaysian archipelago.* (2nd edition.) Honolulu: University of Hawaii Press.

Blevins, Juliette. 2007. A long lost sister of Proto-Austronesian? Proto-Ongan, mother of Jarawe and Ongan in the Andaman islands. *Oceanic Linguistics* 46(1).154–198.

—— To appear. Linguistic clues to Andamanese pre-history: understanding the North-South divide. In Tom Güldemann, Patrick McConvell and Richard Rhodes, eds *Hunter-gatherers and linguistic history: a global perspective.* Cambridge: Cambridge University Press.

Campbell, Lyle. 1988. Review of *Language in the Americas* (Greenberg 1987). *Language* 64.591–615.

—— 1997. *American Indian languages: the historical linguistics of native America.* Oxford: Oxford University Press.

Campbell, Lyle and William J. Poser. 2008. *Language classification, history and method.* Cambridge: Cambridge University Press.

Capell, Arthur. 1948–49. Distribution of the languages in the Central Highlands, New Guinea. *Oceania* 19(2–4).114–129, 234–253, 349–377.

Cavalli-Sforza, Luigi L. 2000. *Genes, peoples and languages.* San Francisco: North Point Press.

Cosgrove, Richard, Jim Allen and Brendan Marshall. 1990. Palaeo-ecology and Pleistocene human occupation in south central Tasmania. *Antiquity* 64(242).59–78.

Cowan, H.J.K. 1957. A large Papuan language phylum in West New Guinea. *Oceania* 28(2).159–166.

—— 1960. Nadere gegevens betreffende de verbreidung der West-Papoease taalgroep (Vogelkop, Nieuw-Guinea). *Bijdragen tot de Taal-, Land- en Volkenkunden* 116(3).350-364.

Croft, William, ed. 2005. Editor's introduction. In Greenberg 2005, x–xxxvii.

Crowley Terry and R.M.W. Dixon. 1981. Tasmanian. In R.M.W. Dixon and B. Blake, eds *Handbook of Australian languages,* vol. 2, 394–421. Canberra: The Australian National University Press.

Dunn, Michael, Ger Reesink and Angela Terrill. 2002. The East Papuan languages: a preliminary typological appraisal. *Oceanic Linguistics* 41.28–62.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley and Stephem C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309.2072–2075.

Dyen I., A.T. James and J.W.L. Cole. 1967. Language divergence and estimated retention rate. *Language* 43(1).150–171.

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

—— 2000. The languages of New Guinea. *Annual Review of Anthropology* 357–404.

—— 2005. Linguistic prehistory in the Sepik-Ramu basin. In Pawley et al., eds, 2005, 109–144.

Franklin, Karl. 1973. Introduction. In K. Franklin, ed. *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*. Canberra: Pacific Linguistics.

Friedlaender, Jonathan, ed. 2007. *Population genetics, linguistics, and culture history in the southwest Pacific: a synthesis*. New York: Oxford University Press.

Friedlaender, Jonathan S. Françoise R. Friedlaender, Floyd A. Reed, Kenneth K. Kidd, Judith R. Kidd, Geoffrey K. Chambers, Rodney A. Lea, Jun-Hun Loo, George Koki, Jason A. Hodgson, D. Andrew Merriwether, and James L. Weber. 2008. The genetic structure of Pacific Islanders. *PLoS Genetics* 4(1).e19.

Gordon, Matthew. 1995. The phonological composition of personal pronouns: implications for genetic hypotheses. In Jocellyn Ahlers, L. Bilmes, J. Guenter, B. Kaiser and Ju Namkung, eds *Proceedings of the 21st annual meeting of the Berkeley Linguistic Society*, 117–128. Berkeley: Berkeley Linguistic Society.

Green, Roger C. 2003. The Lapita horizon and traditions: signature for one set of Oceanic migrations. In C. Sand, ed. *Pacific archaeology: assessments and prospects,* 95–129. Les Cahiers de l'Archéologie en Nouvelle-Calédonie, vol. 15. Noumea: New Caledonia Museum.

Greenberg, Joseph H. 1958. Report on the classification of the non-Austronesian languages of the Pacific. Mimeo. [Report to the Wenner–Gren Foundation for Anthropological Research.]

—— 1960. Indo-Pacific etymologies. Mimeo.

—— 1971. The Indo-Pacific hypothesis. In Thomas A. Sebeok, ed. *Current trends in linguistics 8: linguistics in Oceania,* 807–871. The Hague: Mouton.

—— 1987. *Language in the Americas.* Stanford: Stanford University Press.

—— 2005. *Genetic linguistics. Essays on theory and method.* (ed. William Croft.) Oxford: Oxford University Press.

Groube, L.M., J. Chappell, J. Muke and D. Price. 1986. A 40,000 year old human occupation site at Huon Peninsula, Papua New Guinea. *Nature* 324.453–455.

Haiman, John. 1979. Review of Wurm, ed. 1975. *Language* 55(4).894–903.

Heeschen, Volker. 1978. Review of Wurm, ed. 1975. *Anthropos* 72.978–980.

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.

Kruskal, B., I. Dyen and P.D. Black, 1971. The vocabulary method of reconstructing family trees: innovations and large scale applications. In F.R. Hodson, D.G. Kendall, and P. Tautu, eds *Mathematics in the archaeological and historical sciences*, 30–55. Edinburgh: Edinburgh Press.

Lang, Ranier. 1976. Review of Wurm, ed. 1975. *Kivung, Journal of the Linguistic Society of Papua New Guinea* 9(1).72–80.

Laycock, Donald C. 1975a. A hundred years of Papuan linguistic research: Eastern New Guinea. In Wurm, ed., 1975, 43–115.

—— 1975b. The Torricelli Phylum. In Wurm, ed., 1975, 767–780.

Laycock, Donald C. and John Z'graggen. 1975. The Sepik-Ramu Phylum. In Wurm, ed., 1975, 731–763.

Leavesley, Matthew and John Chappell. 2004. Buang Merabak: additional early radiocarbon evidence for the colonisation of the Bismarck Archipelago, Papua New Guinea. *Antiquity* 78(301).

Loukotka. 1957. Classification des languages Papoues. *Lingua Posnaniensis* (Poznan). 19–83.

Manning, Patrick. 2006. *Homo sapiens* populates the Earth: a provisional synthesis, privileging linguistic data. *Journal of World History* 17(2).115–158.

McElhanon, Kenneth A. 1967. Preliminary observations on Huon Pensinsula languages. *Oceanic Linguistics* 6.1–45.

—— 1970. Lexicostatistics and the classification of the Huon Peninsula languages. *Oceania* 40(3).214–231.

McElhanon, K. and C.L. Voorhoeve. 1970. *The Trans-New Guinea Phylum: explorations in deep-level genetic relationships.* Canberra: Pacific Linguistics.

Mellars, Paul. 2006a. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *PNAS* 103(25).9381–9388.

—— 2006b. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313, no. 5788.796–800.

Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.

Mulvaney, John and Johan Kamminga. 1999. *Prehistory of Australia*. St Leonards, NSW: Allen and Unwin.

Nichols, Johanna and David A. Peterson. 1996. The Amerind personal pronouns. *Language* 72.336–371.

O'Connell, J.F. and J. Allen. 2004. Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *Journal of Archaeological Science* 31.835–853.

O'Connor, Sue. 2007. New evidence from East Timor contributes to our understanding of earliest modern human colonisation east of the Sunda Shelf. *Antiquity* 81.523–535.

Pagel, Mark, 2000. Maximum likelihood models for glottochronology and for reconstructing linguistic phylogenies. In Colin Renfrew, April MacMahon, Larry Trask, eds *Time-depth in historical linguistics,* 189–207. Cambridge: The McDonald Institute of Archaeology.

Pagel, Mark, Quentin Atkinson and Andrew Meade. 2007. Frequency of word use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449.717–720.

Pagel, Mark and Andrew Meade. 2006. Estimating rates of lexical replacement on phylogenetic trees of languages. In Peter Forster and Colin Renfew, eds *Phylogenetic methods and the prehistory of languages*, 173–182. Cambridge: McDonald Institute.

Pavlides, Christina and Chris Gosden. 1994. 35000 year-old sites in the rainforests of West New Britain, Papua New Guinea. *Antiquity* 68.604–610.

Pawley, Andrew. 1998. The Trans New Guinea Phylum hypothesis: a reassessment. In Jelle Miedema, Cecilia Ode and Rien A.C. Dam, eds *Perspectives on the Bird's Head of Irian Jaya, Indonesia,* 655–689. Amsterdam: Editions Rodopi.

—— 2001. The Proto Trans New Guinea obstruents: arguments from top-down reconstruction. In Pawley, Ross and Tryon, eds, 2001, 261–300.

—— 2002. The Austronesian dispersal: languages, technologies, peoples. In Peter Bellwood and Colin Renfrew, eds *Examining the farmingl/language dispersal hypothesis*, 251–273. Cambridge: McDonald Institute of Archaeological Research, Cambridge University.

—— 2005a. The chequered career of the Trans New Guinea hypothesis: recent research and its implications. In Pawley et al., eds, 2005, 67–107.

—— 2005b. Papuan languages. In K. Brown and Anne Anderson, eds *Encyclopaedia of language and linguistics* (2nd edition*)*, vol. 9, 162–171. Oxford: Elsevier.

—— 2007. Recent research on the historical relationships of the Papuan languages, or, What does linguistics say about the prehistory of Melanesia? In Friedlaender, ed., 2007, 36–58.

—— n.d. Some Trans New Guinea Phylum cognate sets. Unpublished Manuscript. Canberra: The Australian National University.

Pawley, Andrew, Robert Attenborough, Jack Golson and Robin Hide, eds. 2005. *Papuan pasts: cultural, linguistic and biological histories of the Papuan-speaking peoples*. Canberra: Pacific Linguistics.

Pawley, Andrew, Malcolm Ross, and Darrell Tryon, eds. 2001. *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton.* Canberra: Pacific Linguistics.

Ray, Sidney H. 1919. A new linguistic family. *The Bible in the World.* October.

Reesink G. 2005. West Papuan languages: roots and development. In Pawley et al., eds, 2005, 185–219.

Renfrew, Colin, April McMahon and Larry Trask, eds. 1999. *Time depth in historical linguistics*. Cambridge: MacDonald Institute for Archaeological Research.

Rhodes, Richard. 1997. On pronominal systems. In Iren Hegedüs, P. Michalove and A. Manster Ramer, eds *Indo-European, Nostratic and beyond, festschrift for Vitaliy V. Shevoroskin*, 293–319. J. Indo-European Studies monograph 22. Washington, D.C.: Institute for the Study of Man.

Ringe, Don. 1992. *Calculating the factor of chance in language comparison.* Philadelphia: American Philosophical Society.

—— 1996. The mathematics of 'Amerind'. *Diachronica* 13(1).135–154.

—— 1999. How hard is it to match CVC-roots? *Transactions of the Philological Society* 97(2).213–244.

Ross, Malcolm D. 1995. The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world: liber amicorum Bert Voorhoeve,* 139–168. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

—— 2000. Pronouns as preliminary evidence for grouping Papuan languages. Unpublished manuscript. Canberra: The Australian National University.

—— 2001. Is there an East Papuan phylum? Evidence from pronouns. In Pawley, Ross and Tryon, eds, 2001, 301–321.

—— 2005. Pronouns as preliminary evidence for grouping Papuan languages. In Pawley, et al., eds, 2005, 15–65.

Ross, Malcolm and Åshild Næss. 2007. An Oceanic origin for Aiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46(2).456–498.

Ruhlen, M. 1991. *A guide to the world's languages.* vol. 1. *Language classification.* (2nd edition) Stanford: Stanford University Press.

Schmidt, Wilhelm. 1952. *Die tasmanischen Sprachen*. Utrecht-Anvers: Spectrum.

Specht, Jim. 2005. Revisiting the Bismarcks: some alternative views. In Pawley et al., eds, 2005, 235–288.

Spriggs, Matthew. 1997. *The Island Melanesians.* Oxford: Blackwell.

Thangaraj, K., G. Chaubey, A.G. Reddy, V.K. Singh and L. Singh. 2006. Unique origin of Andaman Islanders: insight from autosomal loci. *Journal Human Genetics* 51(9).800–804.

Thangaraj, Kumarasamy, Lalji Singh, Alla G. Reddy, V.Raghavendra Rao, Subhash C. Sehgal, Peter A. Underhill, Melanie Pierson, Ian G. Frame and Erika Hagelberg. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Current Biology* 13(2).86-93.

Torrence, Robin, Vince Neall, Trudy Doelman, Ed Rhodes, Chris McKee, Hugh Davis, Roberto Bonetti, Alessandra Guglielmetti, Alberto Manzoni, Massimo Oddone, Jeff Parr and Wallace Cleland. 2004. Pleistocene colonisation of the Bismarck Archipelago: new evidence from New Britain. *Archaeology in Oceania* 39(3).101–130.

Viberg, Åke. 1984. The verbs of perception: a typological study. In Brian Butterworth, Bernard Comrie and Östen Dahl, eds *Explanations for language universals*, 123–262. Berlin: Mouton de Gruyter.

Voegelin, Carl F. and Florence M. Voegelin. 1965. Languages of the World. Indo-Pacific fascicle 5. *Anthropological Linguistics* 7(9).1–114.

Voorhoeve, C.L. 1968. The Central and South New Guinea Phylum: a report on the language situation in south New Guinea. In *Papers in New Guinea Linguistics 8*, 1–17. Canberra: Pacific Linguistics.

—— 1975. Central and western Trans-New Guinea Phylum languages. In Wurm, ed., 1975, 345–459.

—— 2005. Asmat-Kamoro, Awyu-Dumut and Ok: an enquiry into their linguistic relationships. In Pawley et al., eds, 2005, 45–166.

Wickler, S. and M. Spriggs. 1988. Pleistocene occupation of the Solomon Islands, Melanesia. *Antiquity* 62.703–706.

Whitehouse, Paul, T. Usher, M. Ruhlen and W.S-Y Wang. 2004. Kusunda: an Indo-Pacific language in Nepal. *PNAS* 101(15).5692–5.

Wurm, Stephen A. 1954. Notes on the structural affinities of non-Melanesian (Papuan) languages. In W. Koppers, ed. *Die Wiener Schule der Voelkerkunde, Festschrift zum 25 jahrigen Bestand 1929-1954*, 467–472. Vienna: Herold.

—— 1960. The changing linguistic picture of New Guinea. *Oceania* 31(2).121–136.

—— 1964. Australian New Guinea Highland languages and the distribution of their typological features. *American Anthropologist* 66(4-part 2).79–97.

—— 1965. Recent comparative and typological studies in Papuan languages of Australian New Guinea. *Lingua* 15.373–399.

—— 1971. The Papuan linguistic situation. In T.E. Sebeok, ed. *Current trends in linguistics* vol. 8. *Oceania*, 541–657. The Hague: Mouton.

—— 1975a. Possible wider connections of Papuan languages: Papuan and Australian; Greenberg's hypothesis. In Wurm, ed., 1975, 925–932.

—— 1975b. Personal pronouns. In Wurm, ed., 1975, 191–217.

—— 1982. *Papuan languages of Oceania.* Ars Linguistica 7. Tubingen: Gunter Narr.

Wurm, Stephen A., C.L. Voorhoeve and K. McElhanon. 1975. The Trans New Guinea Phylum in general. In Wurm, ed., 1975, 299–322.

Wurm, Stephen A., ed. 1975. *New Guinea area languages.* vol. 1. *Papuan languages and the New Guinea linguistic scene.* Canberra: Pacific Linguistics.

Wurm, Stephen A. and Shiro Hattori, 1981–83. *Language atlas of the Pacific area* (2 vols). Canberra: Australian Academy for the Humanities in collaboration with the Japanese Academy.

Z'graggen, John. 1975. The Madang-Adelbert Range sub-phyum. In Wurm, ed., 1975, 569–612.

9 *A connection between Bird's Head and (Proto) Oceanic*

GER REESINK

1 Introduction

On linguistic and archaeological evidence it is well-established that the Oceanic branch of the Austronesian (An) family had the Bismarck Archipelago as its homeland (Ross 1988; Lynch et al. 2002; Pawley 2003, 2008; Summerhayes 2007). The first speakers of Proto Oceanic (POc) arrived in the Bismarcks from the area of Cenderawasih Bay. Its immediate sister-clade in the An tree is formed by the languages of the South Halmahera West New Guinea (SHWNG) family.

Himmelmann (2005) has shown that the languages of the SHWNG group belong to a typologically aberrant group vis-à-vis the rest of the western An languages. Klamer et al. (2008) argued that it is precisely these differing typological features of what Himmelmann termed the *preposed possessor languages* that are due to contact with the Papuan languages of the *East Nusantara* area, which includes the Bird's Head and Cenderawasih Bay.

Recently, Ross and Næss (2007) and Næss and Boerger (2008) have convincingly shown that the Reefs-Santa Cruz languages are not Papuan, but on the contrary together with other Temotu languages most likely form a first-order subgroup of Oceanic.

As Malcolm Ross has suggested in various places (e.g. Ross 2001:138), some innovations in Oceanic may be due to contact with speakers of Papuan languages. He notes the distinction between alienable and inalienable possession as an example, and I would add the clause-final position of the negative (ad)verb in a number of Oceanic languages.

Continuing the quest to unravel the interaction between the relatively recent An languages and the more ancient languages of various Papuan lineages, in this paper I investigate some possible connections between the languages spoken in the area of POc's precursor and, as we now assume, one of its primary descendants in the Temotu province of the Solomon Islands and north Vanuatu.

In §2 I first give some evidence of apparent (pre-)POc loans in some Papuan languages of the Bird's Head. This is followed (§3.1) by a more specific overview of form and position of some aspectual adverbs in both Papuan and Austronesian languages of eastern Indonesia. In §3.2 I present evidence of the clause-final position of perspectival aspect in Oceanic languages, with a focus on both form and function in languages of the Temotu

province. The specific hypothesis I want to advance is that the clause-final position of perspectival aspect adverbs in Oceanic languages is due to contact with Papuan languages, and that even the form of such adverbs in languages more than 4000 kilometers away from the Bird's Head may be more than an accidental similarity. In the conclusion I will suggest that this enormous geographical distance is bridged by a relatively fast migration of some of the earliest Oceanic speakers.

2 Lexical correspondences Bird's Head and Oceanic

In Reesink (1998) I offered a list of recent and some apparently older An etyma in the Papuan isolate Hatam. The recent ones were clearly due to missionary activities since the middle of the nineteenth century which led to the spread of Biak words in languages all over the Bird's Head (BH), including Hatam. Some other etyma seem to be older, exhibiting regular sound correspondences reflecting both Biak and Hatam phonological systems, for example, Biak *kukər* and Hatam *tut* '(along) with'. Biak is a SHWNG language which shows a regular sound change *t > *k* based on correspondences with other SHWNG languages (and reconstructed protolanguages), as shown for example by *kans* 'weep', reflecting PAn *tangis (Van den Heuvel 2006:24).

In addition to the few items of this earlier list, I will give a few more items in Hatam and some in other BH languages that suggest rather ancient Papuan-An contact, not just on the smaller islands surrounding the Bird's Head, but also on the peninsula itself. In Table 1 PAn or POc reconstructions, as far as I have been able to find them in Wurm and Wilson (1975), Ross (1988) or Ross, Pawley and Osmond (1998), are compared with forms in Biak and Ma'ya of the SHWNG group and some of the Papuan languages of the Bird's Head, Hatam and Mpur.

Table 1: Lexical correspondences between BH Austronesian and Papuan neighbours

	PAn/Pre-POc	POc	Biak	Hatam	Mpur	Ma'ya
banana	*punti	*pudi	(im)byef	wiT		
causative	*pa-	*pa[ka]-	fa-		fa-	
die	*matay	*mate	mai	mai		
ear	*taliŋa		kna	[təŋou]		
embarrassed	*malu		[ma]	mmai		
four	*empat	*pati	fyak	(bə-)tai	bwat	
give	*bəɣaj (?)		[βuk]		[bot]	
hit	*buNuq	*punu(q)		bui[a]		
how many		*pican	fis	pig [pɸiç]		
hundred	*utu-ni		utin	untin	untin	
(loin)cloth	*maru	*malo[b]	mar	mai		
person	*tau[mataq]			tuŋwatu		
inhabited area	*banua	*panua[c]	mnu	[mənu]		
sago	*Rampia	*Rabia			bi[low]	bi[high]
stretch out hand			su	su[d]		

[a] Hatam *bui* means 'hit with hand'.
[b] See Osmond and Ross (1998:98).
[c] POc as given in Green and Pawley (1998:62); the forms in Biak and Hatam refer to 'village'.
[d] In Moi of the west Bird's Head the form is *suu* for this meaning.

Earlier I have speculated whether the initial, syllabic, nasal on many Hatam (intransitive) verbs corresponds to the Austronesian intransitivizing prefix *ma-* (Reesink 2002a). This prefix is clearly reflected in Biak (Van den Heuvel 2006:172) and Taba (Bowden 2001:50), although in neither of these SHWNG languages is it still productive. It seems likely that the geminate /mm/ in *mmai* 'be embarrassed' in Hatam is due to the fossilisation of such a prefix. A number of other intransitive (uncontrolled) verbs that have this particular phonological form, such as *nggieP* 'be unripe', and forms that do not have a counterpart without the initial nasal, such as *nggum* 'be hungry', may also comprise a fossilised *m- prefix. As noted in Reesink (2002a:21) a few pairs of verbs in Hatam with and without an initial nasal may have similar origins, for example, *kwei* 'come' and *ngkwei* 'return'; *kes* 'drop' and *ngges* 'drop (unintentionally)', as explained by one informant. Without further examples, I stated that for many more verbs no such semantic correlation could be given. For other verbs, I still cannot see a clear semantic relation between etyma with an initial nasal and similar forms without it, e.g. *buT* 'peel bark' and *mbuT* 'walk' are two very different events, not related through an intransitive prefix *m-.

Table 2: Examples of Hatam verbs with and without an initial nasal

brei	openly, public	*mbrei*	foreign(er)
bon	make, do	*mbon*	urinate
buT	peel bark of tree	*mbuT*	walk
keK	play	*ngkeK*	skin
kieP	baby of eight months	*nggieP*	unripe
him	very, exact	*nghim*	back (of human, animal, house)
		nggum	be hungry
		nggobiau	be thirsty
		nggwen	be sick
		nggin	cough
		mbraP	speak

The evidence provided by these lexical data is not overwhelming, but it does suggest some ancient contact between Papuan languages of the Bird's Head and surrounding An languages. The An elements in Hatam may well have entered the language via its close relative Mansim, now extinct, that was spoken closer to the coast, near present-day Manokwari.

In addition to the few items in Table 1, elsewhere (Reesink 2002a:298) I have suggested that the negative adverb in Hatam *big* [biç] is cognate with [bar] in extinct Mansim, with corresponding *ba ~ βa ~ (u)wa* in a number of Papuan and Austronesian languages in eastern Indonesia. Not only are the forms cognate but in most of these languages the negative adverb is also found in clause-final position. These East-Indonesian areal features appear to extend as far as Island Melanesia. Forms found in Moluccan languages are presented as external evidence for reconstructed negators in POc (Lynch et al. 2002:88). Having presented their reconstructed forms, Lynch et al. state that 'POc was probably a verb-initial language [.... and] expect a grammaticised negator to be clause-initial.' And indeed this is found in a number of languages, but there are also a large number of languages with a clause-final negator. This order is found, together with reflexes of **bwal(li)*, in languages of the Admiralties, a first-order subgroup of Oceanic.

Such correspondence in both form and (typologically unusual) position suggests more than an independent, parallel development. The clause-final position and even the form in some languages of certain aspectual adverbs result from the same scenario of contact and inheritance.

3 Perspectival aspect marking

All Papuan languages of the Bird's Head have not only a clause-final position for the negator, but also for adverbial elements that express an aspectual notion, which I have labelled 'phasal', when listing a number of non-cognate forms translatable as 'already' (Reesink 1998:617). In fact, the label 'phasal' is somewhat unfortunate. Dik (1997:221) distinguishes four subareas of aspectuality: (i) the type of State-of-Affairs (SoA), also known as *Aktionsart*; (ii) PERFECTIVITY versus IMPERFECTIVITY, relating to the question whether a SoA is presented from an outside point of view as one complete, indivisible whole, or from an inside point of view, as being non-complete or in progress; (iii) PHASAL, specifying the development of a SoA, in terms of beginning (Ingressive), continuation (Progressive) or end (Egressive); and (iv) PERSPECTIVAL, relating the SoA to an outside temporal reference point: Prospective, Immediate Prospective, Recent Perfect, and Perfect.

In this paper I mainly focus on adverbs (some sources call these elements 'particles', but I don't think such a vague category is necessary) that express the notion translatable as 'already'. In some languages this can be contrasted with its antonym 'still, yet'. As Dik notes (1997:240):

> Perspectival aspect can take Phasal Aspects in their scope, as in *John has been crying*, [which conveys] 'John has been in the process of crying in the past, and this is still relevant at the present moment'.

The perspectival aspect ALREADY can indeed have other aspects (or tense or mood) in its scope: *he was already eating, when I came in*; *she had already finished eating*, *I will already have left*, and so on.

The perspectival aspect ALREADY can be characterised as follows:

(1) The SoA holds at a time earlier than the Speaker expected or earlier than the Speaker assumes the Addressee expected.

While (1) places (perhaps undue) emphasis on the Speaker's attitude with respect to the SoA reported, a rather natural corollary is that a SoA must have obtained before a second SoA is related to it: [X already being the case, then Y] is equivalent to [After X, Y]. Some descriptions may lack the actual gloss 'already', but use glosses such as 'perfective' or 'completive' for what seem to be functionally equivalent elements. For this preliminary survey I take some liberty in identifying adverbial elements conveying the perspectival aspect ALREADY, sometimes based on free translations or comments in the text rather than the interlinear gloss provided. Although I do not claim that all languages have lexical (or morphological) material to signal perspectival aspect, a great number of Papuan languages of eastern Indonesia and Austronesian languages of this area and also some languages in Island Melanesia have adverbs in clause-final position that do.

Let us now consider the parallel constructions involving this perspectival aspect in some Papuan and Austronesian languages of eastern Indonesia (§3.1), before we turn to the Oceanic languages of Island Melanesia (§3.2).

3.1 ALREADY in Bird's Head Papuan and neighbouring SHWNG

As representative of the Papuan languages of the Bird's Head, here are some examples from Hatam, where we find *tu* (with its allomorph *su* following the phonetic sequence [-iç]) translatable by 'already' and *yo* 'still, yet'. The form *tu* is illustrated in (2) and the contrast between *tu* and *yo* in (3) and (4), showing that these perspectival aspect markers may occur following the clause-final negator.

HATAM

(2) *Yoni i-kwei leu Mar tu lene ni-ha-gom n-ug ei ...*
they 3PL-come from Warmare already then 1EXC-be-one 1EXC-go LOC
'After they had come from Warmare we all went to ...' (Reesink 1999:90)

(3) *Yoni i-kwei ei Mar big=yo*
they 3PL-come LOC Warmare not=yet
'They have not yet come to Warmare.' (Reesink 1999:96)

(4) *Nab pi-ma binmai big tu*
pig ANAPH-that move not already
'That pig no longer moved.' (Reesink 1999:97; with note regarding the unexplained lack of assimilation to [su])

The similarities in the SHWNG language Taba are striking. Bowden (2001:332) recognizes a particle *do* which he identifies as REALIS mood. Interestingly, he states explicitly that 'realis mood is not obligatorily marked', and the free translations for the examples that contain this particle all use the English adverb 'already', as in (5).

TABA

(5) *Bo-bo-ak-no, dukon Taba ha=wal pa sio do.*
formerly-formerly-to-here eruption Makian CLASS=eight or nine REAL
'From way back in the past up to that time, Makian has erupted eight or nine times already.' (Bowden 2001:332)

As in Hatam (4), *do* 'realis' in Taba may follow the clause-final negator to express 'not anymore', as in (6).

TABA

(6) *Mai a-ne l=pe-ik saguer te-do.*
but DEM-PROX 3PL-make-APPL palm.wine NEG-REAL
'But here they don't make palm wine anymore.' (Bowden 2001:338)

Thus it seems that the clause-final adverb *do* is on a par with what Bowden identifies as the continuative aspect *hu* translated as 'still', which also occupies the clause-final position (2001:333), and both can be characterised as conveying perspectival aspect. A rather surprising fact in Taba, though, is that both perspectivals can occur together; (7) 'was uttered in response to another person's bemoaning the fact that someone who was expected to visit had not arrived' (Bowden 2001:334).

TABA

(7) *I n=wom do hu.*
3SG 3SG=come REAL CONT
'He's still coming.' (Bowden 2001:334)

The form *do* seems to have cognates in other languages of the SHWNG group. For two languages of Cenderawasih Bay, Serui-Laut and Wandamen, the adverb 'already' is given as *to* by Anceaux (1961:71), who also lists *tawa* for Wandamen, and *kwar* for Biak, both also translatable by 'old' for objects.

Since in Biak, as we have seen, *t and *k have merged, the adverb *kwar* appears to reflect the expected cognate form *ku* with an extra element *-ar*, and its function and syntactic position, as described by Van den Heuvel (2006:133), agree with what we find in Taba and (Papuan) Hatam, as illustrated in (8).

BIAK

(8) *Su-mrán kám voi si-wofr afr kwar?*
3DL-walk all but 3PL.AN-blow lime already
'The two walk together, but have they blown lime yet?' (Van den Heuvel 2006:133)

The question in (8) refers to the traditional wedding ceremony in which some people blow lime over the heads of the new couple; in other words, have they already been officially married?

The Biak equivalent for 'still' is the reduplicated form *kaker* of the adverb *ker* which expresses 'continually'. While *ker* occurs post-verbally and thus can intervene between the verbal predicate and any object (if present), *kaker* is only attested in clause-final position (Van den Heuvel 2006:132).

In Tetun of Timor (Van Klinken 1999:235) most examples containing *tiʔan, tiʔa* and *ta* all glossed as 'already' have this adverb in clause-final position. The same observation holds for Leti (Van Engelenhoven 1995:149; 207–208) which has the aspectual adverbs *sala~salmèka* 'already' and *ma:ta* 'still' and *rraa* 'again'.

Finally, Buru, whose genealogical affiliation is not quite clear (Grimes 1991:495–506) has adverbs expressing speaker evaluation, negation, necessity (probability, certainty) in clause-final position (Grimes 1991:232). This includes the adverb *haik*, translated as 'already' and labeled as a post-verbal auxiliary expressing perfective aspect, as illustrated in (9).

BURU

(9) *Da iko gam ak Rana haik.*
[3SG go ALLATIVE up Rana]$_{\text{CLAUSE}}$ [PFV]$_{\text{AUX}}$
'He's gone up to Rana already.' (Grimes 1991:337)

While I have not been able to check all the An languages further to the west, it seems a plausible hypothesis that the clause-final position of the adverb expressing the perspectival aspect ALREADY is found in languages from Timor eastward. To cite just a few sources, in a few languages of Borneo, Mualang (Tjia 2007:194) *(u)dah* 'already' and Begak (Goudswaard 2005:107) *bay* 'already' always take a pre-verbal position. In Kambera (Klamer 1998:120), spoken on Sumba, *mbàda* 'already' is sentence-initial, as with a number of other sentential adverbs. In addition, there are pre- and post-verbal adverbs, but none are exclusively clause-final. For Muna of South-Sulawesi (Van den Berg 1989:188–190) I have not found adequate examples, but focusing adverbs such as *tora* 'again' immediately follow the verb, preceding a post-verbal subject or object.

In other words, the conclusion of this section is that An languages in Maluku have perspectival adverbs in a clause-final position due to their contact with the Papuan

languages of this area. Moreover, there is formal similarity of elements expressing ALREADY between some Papuan and some An languages (Hatam, Biak, Taba).

3.2 ALREADY in Oceanic

In this section I will suggest that the clause-final position of perspectival aspect markers in Oceanic languages is inherited from the Oceanic precursor(s) in eastern Indonesia, perhaps further strengthened by contact with the Papuan languages encountered in the Bismarck Archipelago. Further, that the formal similarity of perspectival adverbs in one of the primary subgroups could be more than accidental.

The reconstructed verb phrase for Proto Oceanic has pre-verbal segments, often portmanteau morphemes indicating person and number of the subject and aspect/mood distinctions (Lynch et al. 2002:85). At the same time, Lynch et al. (2002:46) state that in present-day Oceanic languages 'the final element of a verb phrase is often an aspect morpheme, either enclitic or free.' These observations together suggest that the post-verbal aspect morpheme is of a different nature from the pre-verbal one. And, in fact, what is called the final element of the verb phrase is in many (most?) instances clearly a clause-final adverb, conveying a different aspectuality from the pre-verbal ones.

The clause-final position of such aspectual adverbs in certain languages of Melanesia is quite similar to what we found in both Papuan and An languages of eastern Indonesia, and very likely due to similar contact scenarios. A well-known account of contact-induced levelling of morphosyntactic features is given by Thurston (1982, 1987). He shows that both Papuan Anêm and its Oceanic neighbours, members of the Siassi group such as Lusi and the Bibling languages, such as Lamogai, form a *Sprachbund* quite similar to the Bird's Head and surrounding An languages.

In the West New Britain languages too, we find strictly clause-final adverbs indicating negation and perspectival aspects (Thurston 1987:74–78), as shown for Oceanic Lusi (10) and Papuan Anêm (11).

LUSI

(10) *Paulus i-gali gaea mao.*
Paul SG-spear pig NEG
'Paul didn't spear a pig.' (Thurston 1987:75)

ANÊM

(11) *Me-i axî agonu a de-zik bizaŋ.*
1PL.IRR-arrive at village and 3SG.F.IRR-die already
'By the time we get to the village, she will be dead.' (Thurston 1987:76)

With respect to (11), note that I gloss *bizaŋ* as 'already', although Thurston (1987:76) labels the clause-final adverb as marking *completive* because 'it indicates actions or processes that are complete'. My gloss is motivated by Thurston's observation that the 'Tok Pisin equivalent is *pinis* for which the gloss 'already' is appropriate in most contexts.' He supplied example (11) to show that *completive* does not mark tense, as although most completive events refer to the past, it can also refer to a future event, indicated in (11) by the irrealis verb forms.

Bali-Vitu shows a slightly different picture. This language is classified as an isolate within the Meso-Melanesian linkage of Western Oceanic; while it has undergone the innovations shared by other Meso-Melanesian languages, it has not undergone any of the

innovations which define the subgroups within Meso-Melanesian. (Lynch et al. 2002:102; 362). The language has a number of pre-verbal aspect/mood/sequential markers fused with person, as shown by Ross (2002a:374). Van den Berg and Bachet (2006:98) give IRREALIS *na* for first, *nu* for second, and *ni* for third person, and the PERFECT set *te*, *tu*, and *ti*.

Van den Berg and Bachet (2006:104–108) discuss a few slightly different ways in which the perfect markers can be used: (a) with stative verbs the perfect indicates a situation which is the result of a change that has current relevance; (b) it signals that the action is or was completed at some point of reference; (c) when the action is viewed as located further back in the past; (d) related to the 'anterior' notion, the perfect marker is used in a temporal adverbial clause preceding the main clause; (e) in a temporal clause with time word that follows the main clause; (f) in a main clause indicating a sudden and unexpected action. The (b) usage is frequent in combination with the adverb *kava* 'already'. And this can be found in various positions in the clause: initial or second, preceding the aspect/person marker, as in (12), or clause-finally, as in (13).

BALI-VITU

(12) *Boro kava ti luga kaua..*
pig already PF3 carry dog
'The pig had already carried the dog away.' (Van den Berg and Bachet 2006:106)

(13) *Ia hada kapiru ti mate kava.*
3S see child PF3 die already
'She saw that the child had (already) died.' (Van den Berg and Bachet 2006:106)

As Bali-Vitu shows, a rigidly clause-final position for perspectival aspect markers is certainly not ubiquitous in Oceanic languages, but it is attested in a number of languages, belonging to different branches. Examples include Jabêm of the North-New Guinea linkage, which has, among other clause-final aspectual markers, *su* translated as 'finish' (Ross 2002b:284) or as 'done, having achieved the desired result' (Dempwolff in Bradshaw and Czobor 2005:64), illustrated in (14).

JABÊM

(14) *Òbo gê-gic su.*
cloth 3SG.REALIS-rip done
'The cloth is completely [already?] torn.' (Bradshaw and Czobor 2005:64)

Kokota of the Meso-Melanesian linkage marks TAM categories together with person distinctions by pre-verbal auxiliaries (Palmer 2002:509) and has in addition a second aspect slot in clause-final position following the object, if present. This clause-final slot seems to be reserved for perspectival aspects, for example, *-gu ~u* 'progressive', *fea* 'initially', *nhigo* 'completive'.

Languages of the Papuan Tip subgroup do not seem to have a clause-final slot for such aspectuals, nor for the negator. I could not find any in the sketches for Gapapaiwa or Sudest in Lynch et al. (2002). Papuan Tip languages have in general a clause-initial or pre-verbal negator, as witnessed by Saliba (Margetts 1999:32). Saliba has a verbal suffix *-ko*, which is translated by Saliba speakers as 'already', which may have developed from an earlier complex verb construction with the stem *kohi* 'finish' (Margetts 1999:12–13). Since Saliba, like other Papuan Tip languages, has a clausal OV order this is *ipso facto* clause-final. But in general Papuan Tip languages seem to have (aspectual) adverbs in a pre-verbal position, as illustrated by the Minaveha example (15).

MINAVEHA

(15) *Tau-na ya paisewa avaha Ø-dewa-i..*
one-3SG 3SG.POS work already 3SG-do-OBJ
'He already did his work.' (Lovell 1994: Morphology, example [10])

Thus, perspectival adverbs take a clause-final position in North-New Guinea and Meso-Melanesian linkages, but not in the Papuan Tip linkage. The small sample of languages I have taken to represent these linkages do not allow any lexical reconstruction.

In the introduction I cited the articles by Ross and Næss (2007) and Næss and Boerger (2008) that have provided convincing evidence for the inclusion of the Reefs-Santa Cruz languages in the Oceanic branch of Austronesian. Their suggestion that these languages are part of a primary division within Oceanic may receive further support from the form and function and syntactic behaviour of perspectival aspectual adverbs.

In Äiwoo there are two enclitics marking 'phasal' aspect, as Næss and Boerger call it. The clitic *=to* indicates a transition from one state of affairs to another, while *=jo* focuses on the progression of the event itself (Næss and Boerger 2008:191), illustrated in the examples (16) and (17).

ÄIWOO

(16) *I-lobâku-usi=to.*
PFV-fold-again.TR=PHASAL
'S/he had folded it again (the object was presented to me in a folded state).'

(17) *I-lobâku-usi=jo.*
PFV-fold-again.TR=PHASAL
'S/he folded it again (I saw the act of folding).' (Næss and Boerger 2008:191)

While I have maintained the identical gloss PHASAL for both adverbs as provided by the source cited, it would seem quite feasible to interpret these forms as perspectival, *to* expressing ALREADY and *jo* 'still, ongoing'. And it would seem that a similar perspectival aspect ALREADY is present in languages of the Banks Islands, just south of Temotu.

In Vurës, spoken on Vanua Lava of the Banks Islands, a post-verbal particle *ti* occurs, which according to Catriona Malau is not easily defined in terms of meaning or function. She presented a number of examples (Malau 2007) to investigate whether there is only one word *ti*, or various homophones, and whether it marks aspect or functions as a discourse marker. This form can occur in combination with any one of the preverbal aspect proclitics, *mO=* PFV, *gO=* NON-COMPL, *tO=* PROG, *gOtO=* NEG, and *mitO=* FUT.NEG (the vowels indicated by *o* are underspecified and take their value from the verb stem following), as seen in the two examples (18) and (19). The most parsimonious analysis would be to assume a monosemous form expressing a perspectival aspect, translatable as 'already', and I will gloss the particle as such.

VURËS

(18) *Gö=dun ta i tama-ñ mi=miat ti?*
NON-COMPL=true COMP PERS.ART father-2SG.POS PFV=die already
'Is it true that your father died (and then came back to life)?'
(i.e. had an episode where he was unconscious but was then revived.)

(19) *Qet ni ēl ine ta=van-van ti, nē wo,*
Qet 3SG.UNSPEC see DEM PROG=REDUP-go already 3SG say

"Ei, wona i Dōl ilēkē ta=van-van ti."
hey ANA PERS.ART Dōl DEM.DIST PROG=REDUP-go already
'Qet saw him coming, and he said, "Hey, that's Dōl coming."'

4 Conclusion

Languages can share lexical or morphosyntactic features as a result of three possible scenarios: (i) through inheritance from the same ancestral speech variety; (ii) through contact amongst speakers of distinct languages with linguistic features exchanged in one way or another (borrowing, metatypy, imperfect learning); (iii) through development in two separate speech varieties independently, with the similarities due simply to chance. So the question here is, which of these three scenarios account for the striking similarities in geographically widely separated languages with respect to their perspectival aspect marking?

Firstly, I would say that a construction [V + 'finish'] to indicate some completive aspect is so iconic and cross-linguistically widely attested that independent development is a default case, when the lexical material lacks any indication of cognacy. This may well be the case in Saliba, as suggested above.

Secondly, though, when both form and function as well as syntactic behaviour are part of the correspondence of a linguistic feature in two languages, one of the two other scenarios may have a greater likelihood. My explanation of the aspectual clause-final particles in Äiwoo and Vurës involves both contact and inheritance. The preverbal aspect distinctions in most Oceanic languages may be the result of the complex historical process of reanalysis and fusion suggested by Lynch et al. (2002:85). These distinctions involve a realis versus irrealis mood distinction and further aspectual distinctions relating to the perfectivity/imperfectivity and phasal aspects, discussed by Dik (1997; see above). Whilst the final element of the verb phrase [or clause], which often is an aspect morpheme (Lynch et al. 2002:46), is not mentioned in their reconstruction of POc, it is found in two of the three recognized linkages within the Oceanic subgroup: North New Guinea and Meso-Melanesian. Obviously, the number of languages represented in this study is far too small to make definitive claims, but enough to suggest a plausible hypothesis.

My hypothesis is that clause-final perspectival adverbs, just like clause-final negators (Reesink 2002b), have entered Austronesian languages through contact with Papuan languages in eastern Indonesia **before** the pre-POc speakers left their kin, living near Cenderawasih Bay. This does not deny the possibility that similar configurations could have entered the POc speech community through contact with other Papuan-speaking groups along the north coast of New Guinea or New Britain. If the Temotu languages are indeed a primary subgroup of POc, as Ross and Næss (2007) claim, it is not at all surprising to find an inherited Oceanic trait, which has its origin in Papuan speech communities.

The time frame between the breakup of EMP near Cenderawasih Bay and the arrival of Oceanic speakers in Temotu must have been not more than a few centuries (Summerhayes 2007:21), perhaps about 15 generations or even as few as four generations (Gray et al. 2009:482). Hence contact and inheritance are not at all improbable scenarios to account for convergences found in languages about 4000 kilometers apart.

References

Anceaux, J.C. 1961. *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. 's-Gravenhage: Martinus Nijhoff.

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

Bradshaw, Joel and Francisc Czobor. 2005. *Otto Dempwolff's grammar of the Jabêm language in New Guinea*. Honolulu: University of Hawai'i Press.

Dik, Simon C. (edited by Kees Hengeveld). 1997. *The theory of functional grammar, Part 1: the structure of the clause*. Berlin: Mouton de Gruyter.

Goudswaard, Nelleke. 2005. The Begak (Ida'an) language of Sabah. PhD thesis. Amsterdam: Vrije Universiteit.

Gray, R.D., Drummond, A.J., and Greenhill, S.J. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323.479–483.

Green, Roger and Andrew Pawley. 1998. Architectural forms and settlement patterns. In Ross, Pawley and Osmond, eds, 1998, 67–114.

Grimes, Charles E. 1991. The Buru language of Eastern Indonesia. PhD thesis. Canberra: The Australian National University.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. In K. Alexander Adelaar and Nikolaus. P. Himmelmann, eds *The Austronesian languages of Asia and Madagasca*, 110–181. London: Routledge/Curzon.

Klamer, Marian, 1998. *A grammar of Kambera*. Berlin/New York: Mouton De Gruyter.

Klamer, Marian, Ger Reesink and Miriam Van Staden. 2008. East Nusantara as a linguistic area. In Pieter Muysken, ed. *From linguistic areas to areal linguistics*, 95–149. Amsterdam/Philadelphia: John Benjamins.

Lovell, Larry Lee. 1994. Minaveha grammar: a study of an Austronesian language of the Milne Bay Province, Papua New Guinea. Unpublished manuscript. Ukarumpa: SIL.

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

Malau, Catriona. 2007. The particle *ti* in Vurës, Vanua Lava, Vanuatu. Paper presented at 7th Conference on Oceanic Linguistics, Nouméa, Nouvelle Calédonie.

Margetts, Anna. 1999. *Valence and transitivity in Saliba, an Oceanic language of Papua New Guinea*. Nijmegen: MPI series in Psycholinguistcs.

Næss, Åshild and Brenda H. Boerger. 2008. Reefs-Santa Cruz as Oceanic: evidence from the verb complex. *Oceanic Linguistics* 47.186–213.

Osmond, Meredith and Malcolm Ross. 1998. Household artifacts. In Ross, Pawley and Osmond, eds, 1998, 67–114.

Palmer, Bill. 2002. Kokota. In Lynch, Ross and Crowley, eds, 2002, 498–524.

Pawley, Andrew. 2003. The Austronesian dispersal: languages, technologies, peoples. In Peter Bellwood and Colin Renfrew, eds *Examining the farming/language dispersal hypothesis*, 251–273. Cambridge: McDonald Institute of Archaeological Research, Cambridge University.

—— 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: Lincom Europa.

Reesink, Ger. 1998. The Bird's Head as Sprachbund. In Jelle Miedema, Cecilia Odé and Rien A.C. Dam, eds *Perspectives on the Bird's Head of Irian Jaya, Indonesia*, 603–642. Amsterdam: Rodopi.

—— 1999. *A grammar of Hatam*. Canberra: Pacific Linguistics.

—— 2002a. The Bird's Head languages compared. In Ger Reesink, ed. *Languages of the eastern Bird's Head*, 1–44. Canberra: Pacific Linguistics.

—— 2002b. Clause-final negation: structure and interpretation. *Functions of language* 9.239–268.

Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian Languages of western Melanesia*. Canberra: Pacific Linguistics.

—— 2001. Contact-induced change in Oceanic languages in north-west Melanesia. In Alexandra Y. Aikhenvald and R.M.W. Dixon, eds *Areal diffusion and genetic inheritance*, 134–166. Oxford: Oxford University Press.

—— 2002a. Bali-Vitu. In Lynch, Ross and Crowley, eds, 2002, 362–386.

—— 2002b. Jabêm. In Lynch, Ross and Crowley, eds, 2002, 270–296.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. 1998. *The lexicon of Proto Oceanic,* vol. 1: *material culture*. Canberra: Pacific Linguistics.

Ross, Malcolm, and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46.456–498.

Summerhayes, Glenn R. 2007. Island Melanesian pasts: a view from archeology. In Jonathan Scott Friedlaender, ed. *Genes, language, and culture history in the southwest Pacific*, 10–35. Oxford: Oxford University Press.

Thurston, William R. 1982. *A comparative study in Anêm and Lusi*. Canberra: Pacific Linguistics.

—— 1987. *Processes of change in the languages of north-western New Britain*. Canberra: Pacific Linguistics.

Tjia, Johnny. 2007. A grammar of Mualang: an Ibanic language of western Kalimantan, Indonesia. PhD dissertation. Leiden: Leiden University.

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

Van den Berg, René and Peter Bachet. 2006. *Vitu grammar sketch*. Data papers on Papua New Guinea languages, vol. 51. Ukarumpa: SIL.

Van Engelenhoven, Aone. 1995. A description of the Leti language. PhD thesis. Leiden: Leiden University.

Van den Heuvel, Wilco. 2006. Biak: description of an Austronesian language of Papua. PhD thesis. Amsterdam: Vrije Universiteit.

Van Klinken, Catharina. 1999. *A grammar of the Fehan dialect of Tetun, an Austronesian language of West Timor*. Canberra: Pacific Linguistics.

Wurm, S.A. and B. Wilson. 1975. *English finderlist of reconstructions in Austronesian languages*. Canberra: Pacific Linguistics.

10 *How many branches in a tree? Cua and East (North) Bahnaric*

PAUL SIDWELL

1 Introduction[1]

The Austroasiatic phylum presents a discontinuous patchwork of perhaps 170 languages stretching from India to Indo-China and the Malay Peninsula. Despite being known to linguists (and subjected to significant comparative studies) for more than a century, the internal classification of Austroasiatic remains extremely problematic. Specialists agree on the identification of a dozen branches, yet they find no broad agreement on the details of how these branches relate, or how most of them are internally structured. This is not always apparent to anyone consulting the reference literature, which will typically quote a single authority or simply present an unsubstantiated scheme which one cannot otherwise assess. The persistent researcher will actually find as many classifications as there are scholars offering views on the subject, and a regrettable shortage of survey works or programmatic discussions. The problem is so bad that other disciplines, keen for the historical and cultural inferences that may be drawn from well researched language classifications, throw up their hands in frustration.

> Austroasiatic languages are the most poorly researched of all those under discussion. Many are not documented at all and some recently discovered in China are effectively not classified. […]. Austroasiatic is conventionally divided into two families, Mon-Khmer (in SE Asia) and Muṇḍā (in India). Diffloth (2005:79) now considers Austroasiatic to have three primary branches but no evidence for these realignments has been published. Indeed Austroasiatic classification has been dogged by a failure to publish data, making any evaluation of competing hypotheses by outsiders a merely speculative exercise. (Blench 2008, 117–118)

As Blench mentions, the reference literature conventionally distinguishes Munda and Mon-Khmer families, the latter comprising some 11 branches or the bulk of the phylum. This classification is typological only, and is based principally on the studies of Pinnow (1959, 1963, etc.). Munda languages are polysynthetic and SOV in syntax, while Mon-Khmer are more or less isolating and SVO (especially, for example, Vietnamese). The

[1] Research that made this paper possible was supported by generous assistance of the National Endowment for the Humanities (Washington DC). Any views, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities. I would also like to thank Vurnell Cobbey (SIL, Dallas) for valuable assistance with access to archives and introductions to researchers and SIL Dallas based personnel.

contemporary list of Mon-Khmer branches was established by lexicostatistical studies of the 1960s and 1970s (especially Thomas 1966; Thomas and Headley 1970), with no subsequent analyses contradicting the broad findings of lexicostatistics.

This paper is concerned with the Bahnaric branch of Mon-Khmer languages. Into the 1960s, scholarship (e.g. Lebar et al. 1964) recognised Mon-Khmer as consisting of Mon, Khmer, plus scores of poorly described minor languages stretching up the Mekong River basin into the hills of Vietnam, Laos and Thailand. One of the largest of these in the Vietnam Central Highlands, Bahnar, had been known to the west via catholic missionaries and colonial officials since the 1860s. It, along with an emerging list of closely related tongues, was characterised as *Bahnaran* in English language scholarship. Bahnaran was not really a taxonomic distinction, but a convenient label for languages apparently close to Bahnar, in an otherwise sprawling dialect chain. It was Thomas (1966) who coined the term *Bahnaric*, and established clear lexical criteria for Bahnaric versus other Mon-Khmer branches in Vietnam. His Bahnaric united Bahnar with *Stiengan* (Stieng plus several other languages that had been grouped with Cambodian in the earliest accounts) as North and South Bahnaric sub-branches respectively. Shortly after that Thomas and Headly (1970) examined various wordlists from Laos, and impressionistically identified a third Bahnaric branch, West Bahnaric (see Table 1), thus giving us three branches corresponding neatly to three geographical areas. This effectively established the frame through which subsequent investigations would pose the question of how to classify the Bahnaric languages: i.e. does languages X belong to North, South or West Bahnaric, or do we need to posit new divisions?

Table 1: Bahnaric classification by Thomas and Headley (1970)

West Bahnaric	South Bahnaric	North Bahnaric
1. Loven (Jru)	1. Stieng	1. Bahnar
2. Nyaheun, Prou	2. Central Mnong (Preh, Biat etc.)	2. Rengao
3. Oi, The	3. Southern Mnong (Nong, Prâng)	3. Sedang
4. Laveh	4. Eastern Mnong (Gar, Chil, Kuanh, Rơlơm)	4. Halang
?5. Brao, Krưng, Kravet	7. Kơho	5. Jeh
6. Sok	8. Chrau (Jro)	6. Mơnơm (Bơnâm)
7. Sapuan		7. Kayong (Cagiương)
8. Cheng (Jeng)		8. Hrê (Davak)
		9. Cua (Kor, Traw)
		10. Takua
		11. Tơdrah (Didrah)

As more data came to hand, and preliminary comparative-historical studies were done (e.g Blood 1966; Smith 1972), it became clear that the simple three coordinate sub-branch model was inadequate. Bahnar did not seem to sit so nicely with the other northern languages; a major concern being that Bahnar lacks the contrastive phonation types ('registers') of other northern languages (such as Sedang, Rengao, Jeh), leading Smith (1972) to speculate that it might really be South Bahnaric. Others pointed out that Bahnar seemed to have striking affinities to Alak (in Laos) and Tampuon (in Cambodia), prompting Gregerson et al. (1976), and Thomas (1979), to propose a Central Bahnaric sub-branch to accommodate these.

It had also caught Smith's attention that Cua (also known as Kol, Kor or Traw[2]), although geographically nested among North Bahnaric languages, lacked registers, and was anomalous is various other ways. In particular, Cua (along with neighbouring Kơtua) shows a hardening of final nasals conditioned by the manner of articulation of the prevocalic consonants. This innovation motivated Smith (1973) to suggest an Eastern North Bahnaric clade, apparently a daughter of Proto North Bahnaric, although the precise relation to other North Bahnaric languages was left unclear.

As Figure 1 and Table 2 (below) indicate, Smith's suggestion was taken up generally.

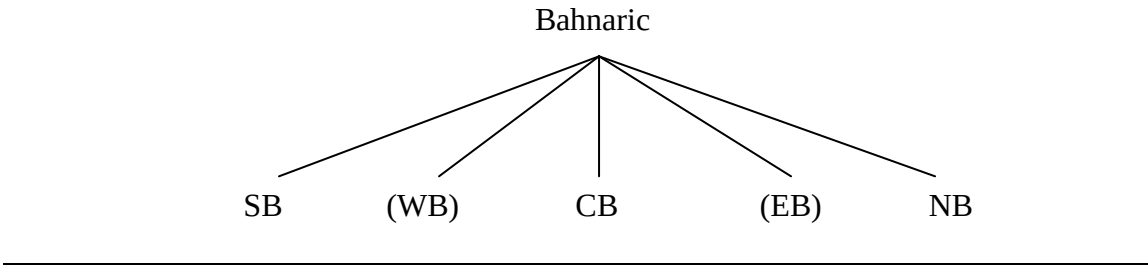


Figure 1: Bahnaric tree by Gregerson et al. (1976:2002)

Table 2: Bahnaric classification by Thomas (1979:183)

North Bahanric:	Sedang, Hrê, Halăng, Jeh, Rengao
South Bahnaric:	Kơho, chrau, Mnong, Stieng
West Bahnaric:	Laven, Nyaheun, Cheng, Oi, Laveh, Brao
Central Bahnaric:	Bahnar, Tampuon, Alak
Eastern Bahnaric:	Cua, Takua (?)[3]

More recently, this writer (Sidwell 2002) proposed a revised Bahnaric classification, with three coordinate sub-branches, North, West and Central, with South subordinated to Central Bahnaric. In that scheme Cua was placed within Central Bahnaric, as an Eastern division.[4]

[2] According to Maier and Burton (1981:2): Cua is a member of the Bahnaric branch of the Mon-Khmer family of languages. The majority of the Cua people, who number 10,000–15,000, live in the mountain area of Tra Bong district in Quang Ngai province, central Vietnam. The Cua people call themselves 'Kool', their term for 'montagnard' in general, this name being modified to 'Cua' by the Vietnamese. The Cua people who live in the Tra Bong valley are called Kool Dong, 'Valley Cua', and the Cua who live in the mountains are called Kool Doot or Kool Taal 'High Cua'.

[3] The listing by Thomas of Takua alongside Cua is incongruous, since it is solidly North Bahnaric. Perhaps Thomas confused Takua with Kơtua.

[4] That study reviewed Bahnaric classification in detail and offered a revised model based upon historical phonology, in contrast to other studies which have applied mainly lexical treatments. I am indebted to one reviewer of the present paper who suggested that I consult specifically Theraphan L-Thongkum (2001) plus an unpublished manuscript by of Gerard Diffloth, which the reviewer states set out 'detailed phonological criteria for subgrouping within […] Bahnaric'. Unfortunately the works I was referred to do no such thing: Theraphan offers some lexical and phonological arguments for a hypothetical 'North West Bahnaric' clade (which I dispute), and Diffloth only points out that reflexes of the 'bone' etymon usefully correlate with Bahnaric sub-groupings. Diffloth's point is correct, but he offers no Bahnaric reconstruction to justify a modelling of branching structure.

However, that classification of Cua as East Central Bahnaric was really a default classification, based mainly on the apparent failure of Cua to show any distinctively Northern or Western innovations. Subsequent analysis of a larger Cua data set now indicates that I misanalysed key sound correspondences, which I had taken as indicative of Cua being Central Bahnaric. The problem is difficult to investigate satisfactorily from the perspective of historical phonology since lexical replacements in Cua have been so extensive it is hard to find multiple examples of etyma which show key sound changes. However, access to a larger dataset has allowed me to make a more comprehensive analysis of Cua historical phonology, and the main result has been to indicate that Cua developed directly and independently from Proto Bahnaric. In that case it seems appropriate to treat Cua as a fourth coordinate branch of Bahnaric, or *East Bahnaric*. Kơtua, on the other hand, classified as a sister of Cua by Smith (1973), is found here to be solidly North Bahnaric, although sharing a common rule governing a hardening of final nasals.

2 Data Sources for Cua

Cua is not well documented by Western scholars, and one is largely restricted to using manuscript materials. Sources readily available in English (via the Summer Institute of Linguistics [SIL] in Dallas and Bangkok) are:

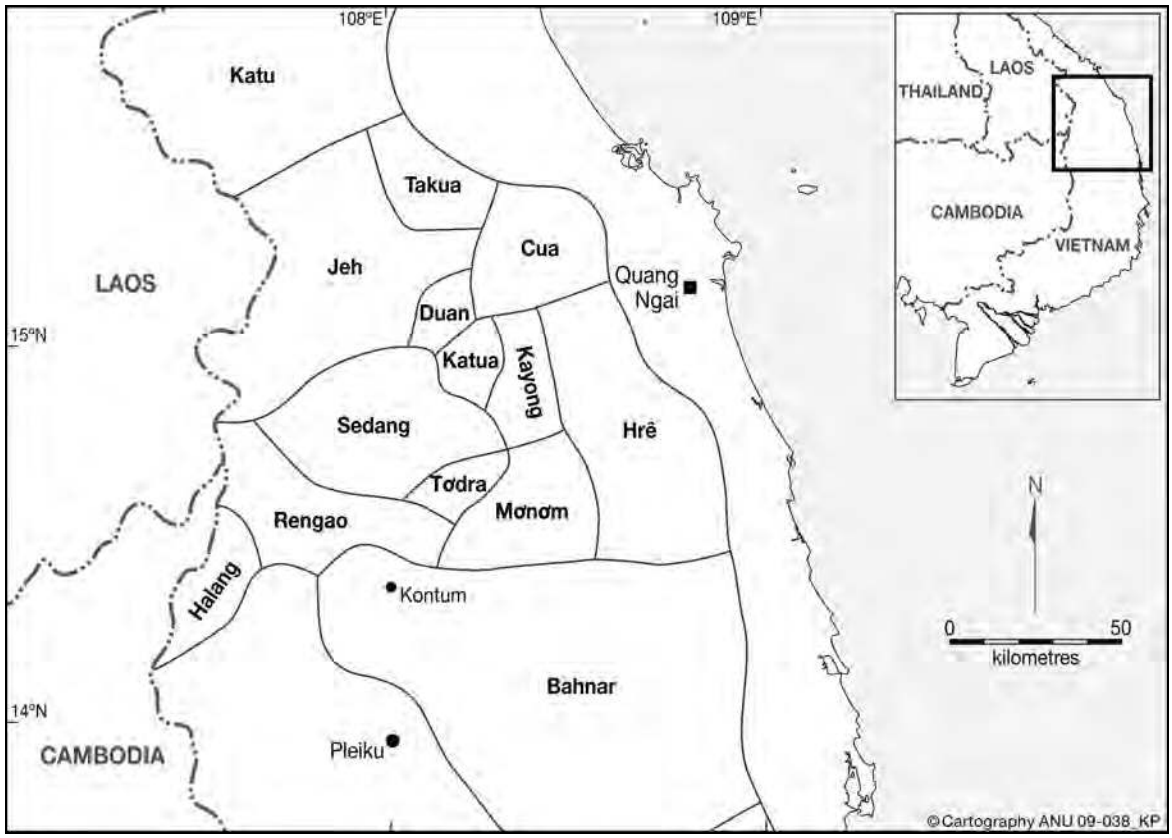
- Phillips (1959), a manuscript lexicon of 187 words in four dialects
- Burton (1969), a discussion of clause structure,
- Maier (1969), an analysis of phonemes,
- Maier and Đình Van Cau (1976) a Cua-Vietnamese-English thesaurus in the form of a computer printout,[5]
- Maier and Burton (1971), a manuscript lexicon of 281 words of a lowland Cua dialect
- Smith (1973), a discussion of classification,
- Maier and Burton (1981), a manuscript rhyming dictionary of somewhat more than 3000 items, annotated to the effect that it is based on a 1966 manuscript

In addition there are some minor literacy and bible translation notes at the David Thomas Library (SIL) in Bangkok, and this writer has collected some very short wordlists in the field.

3 Cua and Kơtua

Smith (1973) tentatively identified Cua and Kơtua (sometimes *Katua*) as East North Bahnaric, that is, an eastern clade of North Bahnaric. One can see from the map reproduced here as Map 1 how these two languages lay in relation to establish North Bahnaric languages (Takua, Jeh, Duan, Kayong, Hrê, Sedang, Rengao, Halang, Tơdra, Mơnơm).

[5] This appears to be the '1700 word dictionary' mentioned by Smith (1973:117) as his principal source.



Map 1: Fragment of SIL map 'Ethno-linguistic minorities of South Vietnam, revised March 1966'.

Smith's analysis found a 'unique but shared phonological shift' (1973:113), in the form of a regular hardening of final nasals to stops, specifically in syllables that do not also have pre-vocalic nasals or laryngeals. This was illustrated by with some 30 lexical comparisons.

Table 3: Examples of nasal hardening rule in Cua and Kơtua[6]

Gloss	Cua	Kơtua	North Bahnaric	Central Bahnaric	West Bahnaric	Proto Bahnaric
bird	*se:p*	*cip*	*cem (Jeh *cim*)	*ce:m (Bahnar *sɛ:m*)	*ce:m (Jru' *ce:m*)	*-m
four	*pɔ:t*	*po:t*	*puan (Jeh *puan*)	*puan (Bahnar *pwan*)	*puan (Jru' *puan*)	*-n
eight	*thə:m*	*naha:m*	*tha:m (Jeh *təha:m*)	*tŋha:m (Bahnar *thŋa:m*)	*tha:m (Jru' *tha:m*)	*-m
night	*kamɨŋ*	*maŋ*	*maŋ (Jeh *maŋ*)	*maŋ (Bahnar *maŋ*)	*maŋ (Jru' *maŋ*)	*-ŋ

[6] West Bahnaric data and reconstructions in this example and henceforth are taken from Sidwell and Jacq (2003), North Bahnaric data and reconstructions are from Sidwell (n.d.) a draft monograph widely distributed for comments in 2002 and still available upon request, and Central Bahnaric and Proto Bahnaric data and reconstructions are from as yet uncirculated manuscripts by Sidwell.

Quite understandably, this fact was interpreted as indicating a sub-grouping within North Bahnaric, for which the term Eastern North Bahnaric (ENB) was proposed.[7] However, it is evident—even based on the brief data provided by Smith (1973)—that there are strong empirical difficulties with this classification. The essential problem is that, despite the sharing of a regular conditioned sound change, only Kơtua exhibits other features which are characteristic of North Bahnaric.

Table 4: Vowel restructuring changes

Gloss	Cua	Kơtua	North Bahnaric	Central Bahnaric	West Bahnaric	Proto Bahnaric
liver	*klə:p*	*kliap*	*klè:m (Rengao *kl̤ẹ:m*)	*klə:m (Bahnar (*klə:m*)	*klə:m (Jru' *klʌ:m*)	*ə:
leg	*ɟo:k*	*nɟe:k*	*è:ŋ (Rengao *ɟẹ:ŋ*)	*ɟɨŋ (Bahnar *ɟəŋ*)	*ɟɨŋ (Jru' *ɟɨŋ*)	*ɨ
breath	*su:m*	*siam*	*ɟəhè:m (Rengao *cəhẹ:m*)	*ɟhɨ:m (Alak *ʤahɯ:m*)	*ɟhə:m (Jru' *cəh:m*)	*ɨ:
bird	*se:p*	*cip*	*cem (Rengao *c̤im*)	*ce:m (Bahnar *sɛ:m*)	*ce:m (Jru' *ceɨ:m*)	*e:
sky	*ple:k*	*plik*	*pliŋ (Rengao *pl̤iŋ*)	*ple:ŋ (Tampuon *plaiŋ*)	*ple:ŋ (Su' *ple:ŋ*)	*e:
mushroom	*phale:t*[8]	–	*psit (Rengao *bətsi̤t*)	–	*pse:t (Jru' *pse:t*)	*e:

The North Bahnaric languages are readily characterised by a restructuring of their vowels which resulted in the emergence of a system of contrastive phonation types or registers; typically breathy versus modal voice. This restructuring included chain shift in which there was a shortening of *e: to /i/ and /e/, and a fronting of *ə: and *ɨ: to /e:/ and /i:/ respectively. In the now otherwise crowded mid-vowel space further mergers were avoided with the emergence of a register distinction in the form breathy versus modal voice. However, it is immediately evident that, while Kơtua reliably shows reflexes consistent with North Bahnaric vowel restructuring, Cua does not (Table 4).

It is also clear that Cua shows some phonological innovations that set it apart from all other Bahnaric sub-groups (see Tables 5 and 6).

[7] Diffloth and Zide (1992) proposed an Eastern division within North Bahnaric that he calls 'Cua-Kayong', and relegates Kơtua to 'uncertain' status. Unfortunately no explanation for this has been published, so it is not viable to critique Diffloth's analysis in this paper. I will say however that it is presently my view (on the basis of comparative reconstruction) that Kayong groups with Jeh-Halăng within North Bahnaric, consistent with the statement of Cooper and Cooper (1966:87) that Kayong is 'mutually intelligible with Halăng', (which cannot be said for Cua).

[8] The regular reflex of *s in Cua is /hl/, but in this case was recoded as /l/, I suspect in error.

Table 5: Raising of *aː and *a to /ɨː/ and /ɨ/ after prevocalic nasals in Cua

Gloss	Cua	North Bahnaric	Central Bahnaric	West Bahnaric	Proto Bahnaric
year	*sanɨm*	*hnam (Jeh *hnam*)	*cnam (Bahnar *sənam*)	– (*kmɔː)	*a
sun/day	*hŋɨj*	– (*hiː)	*tŋaj (Alak *taŋaj*)	*tŋaj (Jru' *taŋaj*)	*a
sweet	*ŋɨːm*	*ʔŋaːm (Jeh *ʔŋaːm*)	*ʔŋaːm (Bahnar *ʔŋaːm*)	*ʔŋaːm (Jru' *ʔŋʌːm*)	*aː
far	*saŋɨːj*	*sʔŋaːj (Jeh *ʔiʔŋaːj*)	*cŋaːj (Alak *caŋaːj*)	*cŋaːj (Jru' *hŋaːj*)	*aː

Table 6: Shift of prevocalic ***s** > /hl/ in Cua (merging with *sl-, and remaining distinct from reflexes of *c which merged with *s in South Bahnaric, Bahnar, and Tampuon)

Gloss	Cua	North Bahnaric	Central Bahnaric	West Bahnaric	Proto Bahnaric
hair	*hloːk*	*sək (Jeh *suk*)	*sɔk (Bahnar *sɔk*)	*sɔk (Jru' *sok*)	*s-
divide/share	*ʔahlɔːk*	*sɔŋ (Halang *sɔaŋ*)	*ʔəsɔːŋ (Bahnar *ʔəsɔːŋ*)	*sɔːŋ (Brao *sɔːŋ*)[9]	*s-
honey(bee)	*hluːt*	*sut (Halang *sut*)	*su(ː)t (Bahnar *suːt*)	*sut (Jru' *suːt*)	*s-
tail	*hlɔːj*	– (Jeh *tɛːŋ*)	– (Bahnar *kjɛŋ*)	*sɔɔj (Jru' *suaj*)	*s-
leaf	*hlaː*	*hlaː (Jeh *laː*)	*hlaː (Bahnar *hlaː*)	*slaː (Jru' *hlaː*)	*sl-
dog	*sɔː*	*cɔː (Jeh *cɔː*)	*cɔː (Bahnar *sɔː*)	*cɔː (Jru' *cɔː*)	*c-
to eat	*saː*	*caː (Jeh *caː*)	*caː (Bahnar *saː*)	*caː (Jru' *caː*)	*c-

These data suggest that, with the exception of the nasal hardening rule, Cua has a unique phonological history which lacks any indication of distinctly North Bahnaric developments. Therefore, any linkage between Cua and North Bahnaric, if it existed, preceded the North Bahnaric vowel restructuring, presumably at a very early stage.

4 Bone of contention

In Sidwell (2002) I proposed phonological criteria for classifying Bahnaric languages into three branches: West, North, and Central (with the latter taking in South Bahnaric, Bahnar, Alak etc.). The first of these criteria is an ɨː : iː distinction which is maintained in West Bahnaric, but lost by a general merger to /iː/ in the rest of the family (with new examples of /ɨː/ emerging secondarily, such as the raising of /aː/ in Cua, discussed above). Examples (noting especially the Cua /ia/ reflexes with final velars and /h/) are given in Table 7.

[9] Brao *sɔːŋ* 'to pay' may be borrowed from/influenced by Khmer *sɔːŋ* 'give back, payback, restore, compensate'.

Table 7: Reflexes of Proto Bahnaric * i: and * ɨ:

Gloss	Cua	North Bahnaric	Central Bahnaric	West Bahnaric	Proto Bahnaric
banana	*pare:t*	*pri:t (Jeh *pr̤iat*)	*pri:t (Bahnar *pri:t*)	*prɨ:t (Jru' *prɨət*)	*prɨ:t
weep/cry	*ɲim*	*ɲi:m (Kayong *ɲiem*)	*ɲi:m (Alak *ɲi:m*)	*ɲɨ:m (Jru' *ɲɨəm*)	*ɲɨ:m
frog	—	*ki:t (Jeh *k̤iat*)	*ki:t (Bahnar *ki:t*)	*ki:t (Jru' *ŋkiət*)	*ki:t
forest	*bari:*	*bri: (Jeh *br̤i:*)	*bri: (Bahnar *bri:*)	*brəj (Jru' *brɛ̆j*)	*bri:
to hoe/field	*ɟiak* 'field'	*ɟi:k (Jeh *ɟ̤iak*)	*ɟi:k (Stieng *ɟik*)	—	*ɟi:k
civet cat	*cpiak*	*spi:k Jeh *sip̤iak*)	*cpi:k (Stieng *pik*)	—	*cpi:k
to dry in sun	*tiak*	*ti:ŋ (Jeh *t̤iaŋ*)	—	*tɨ:ŋ (Jru' *piəs*)	*tɨ:ŋ
citrus/orange	*piah*	—	—	*pɨ:s (Jru' *piəs*)	*pɨ:s

Table 8: Reflexes of Proto Bahnaric *ts, *t and *s

Gloss	Proto Bahnaric	West Bahnaric	Central Bahnaric	North Bahnaric
bone	*ktsɨ̤:ŋ	*k[r]ʔtɨ:ŋ (Jru' *ktɨəʜ*)	*kti:ŋ (Bahnar *kəti:ŋ*)	*ksi:ŋ (Jeh *kəs̤iaŋ*)
tree trunk/base	*tsə:m	*tə:m (Jru' *tʌ:m*)	*tə:m (Bahnar *tə:m*)	*se:m (Halang *sɛɛ̆m*)
monitor lizard	*ktsu:m	—	*ktu:m (Chrau *kətu:m*)	*ksu:m (Jeh *kəsṳam*)
cloud/fog	*tsuk	—	*tuk (Bahnar *tuk*)	*suk (Jeh *kəsṳk*)
arm, hand	*ti:	*tej (Jru' *tɛ̆j*)	*ti: (Bahnar *ti:*)	*ti: (Jeh *t̤i:*)
hot	*[]toʔ	*ʔtoʔ (Jru' *ʔtoʔ*)	*toʔ (Bahnar *tɔʔ*)	*toʔ (Jeh *tuʔ*)
hair	*sɔk	*sɔk (Jru' *sok*)	*sɔk (Bahnar *sɔk*)	*sɔk (Jeh *suk*)
honeybee	*su[:]t	*sut (Jru' *su:t*)	*su:t (Bahnar *su:t*)	*sut (Halang *sut*)

At first glance one might suppose that the sharing of this merger groups all of Central and North Bahnaric and Cua into one clade coordinate with West Bahnaric. However, this cannot be the case; there is an interesting t- : s- correspondence, provisionally reconstructed here as *ts- (consistent with Smith 1972, and reverting from the *sʔ- proposed in Sidwell 2002), which indicates West Bahnaric and Central Bahnaric sharing a merger of *ts- and *t- to /t/, versus a merger of *ts and *s- to /s/ in North Bahnaric.[10] Examples (with last four lines illustrating Proto Bahnaric *t- and *s- for contrast) are given in Table 8.

The *i̱: and *ts correspondence patterns overlap in a manner suggesting a three way dialect split within the protolanguage, ultimately correlating with West, Central and North Bahnaric sub-branches. Remarkably, both of these features occur in the 'bone' etymon, making it a uniquely useful exemplar. This is represented diagrammatically as follows:

West Bahnaric	Central Bahnaric	North Bahnaric
*ts, *t > *t		*ts, *s > *s
*i̱: > *i̱ː *iː > *iː	*i̱:, *i: > *i:	
*k[r]ʔtiːŋ (Jru' *kətɨəŋ*)	*ktiːŋ (Bahnar *kəti:ŋ*)	*ksiːŋ (Jeh *kəsi̱aŋ*)

Somewhat frustratingly, it appears that Cua retains only one *ts- etymon, namely *khiak* 'bone'. In Sidwell (2002) I suggested that /e:/ is the regular Cua reflex of Proto Bahnaric *i̱: and *i: (citing the 'banana' etymon) and on that basis asserted that Cua *kate:k* 'lower back; back rib' is probably the direct reflex of Proto Bahnaric *ktsi̱:ŋ (reconstructed provisionally as *gəsʔi̱:ŋ in 2002). However, the correspondences adduced here clearly indicate that in the environment of a final velar we would expect to see *i̱: reflected as /ia/, confirming the likelihood that *khiak* is the reflex of Proto Bahnaric *ktsi̱:ŋ.

This leaves us with the challenge of explaining the medial /h/ of Cua *khiak* as the reflex of Proto Bahnaric *ts, since we might otherwise expect something more like a [t] or [s]. Significantly, it turns out that in the Phillips (1959) wordlists, four forms for 'bone' are recorded as follows:

Village name	Form	Village name	Form
Cha Lɒm	*khiik*	Na Lai	*kliʌk*
Tra Nom	*khiək*	Dak Kɒmɔy	*kliək*

The variation in these forms between medials /h/ and /l/ strongly suggests an underlying /hl/ which, as shown elsewhere, is the regular reflex of Proto Bahnaric *s. This suggests a pre-Cua *ksiːŋ corresponding nicely to PNB *ksiːŋ 'bone'. Yet it would be far too bold to take this one change, common to North Bahnaric, as indicative. As demonstrated above, Cua independently lenited *c-, *s- and *sl-, so a similarly independent development of *ts- > *s would be quite unmarked, and not particularly indicative of common descent from Proto North Bahnaric. This leaves us with no persuasive historical phonological evidence for anything other than the independence of Cua within Bahnaric.

[10] In fact /ts/ is the normal reflex of *s- in several North Bahnaric languages: Rengao, Kaco', Ramam.

5 Distinctive lexicon?

We are left with the evidence of distinctive vocabulary[11] to test the strength of any claim of special relation to either North or Central Bahnaric. How does such evidence stack up? Thomas (1979) conducted an exercise of testing Bahnaric languages according to a set of 34 categories for which distinctive vocabulary is evident for each of the sub-branches. His results are quoted here:

> These words show the western cluster agreeing fairly often with the northern cluster (9 times) but seldom with the southern cluster (2 times). The agreement between Alak, Tampuan, and Bahnar is striking: 18/31 times they all agree, and once (No.10) they agree with nothing else in Bahnaric supporting them. In contrast, Bahnar and Tampuan agree against Alak only 5 times. The Alak-Tampuan-Bahnar group agree fairly often with the northern and western cluster but seldom with the southern cluster.
>
> Cua shows itself more isolated, frequently having idiosyncratic words. Cua agrees fairly often with the northern and Alak-Tampuan-Bahnar groups, but seldom with the southern and western groups. (Thomas 1979:177–178)

The agreements that Thomas found between Cua and 'the northern and Alak-Tampuan-Bahnar groups' can all be accounted for as lexical retentions found across more than one sub-branch, while the less frequent agreements with 'the southern and western groups' reflect the extent of lexical innovations in the latter. Of the 34 sets, Cua agrees with both North Bahnaric and Bahnar/Tampuon or Alak (representing Central Bahnaric) in all but 11 cases. Of the latter, Cua shows eight idiosyncratic forms (mainly loans), one isogloss with South Bahnaric, and two isoglosses with West Bahnaric. Crucially, neither in Thomas' study, nor in my examination of Maier and Burton's (1981) vocabulary, have I found a single example of a distinctively North Bahnaric lexical innovation that is uniquely shared with Cua. Clearly it is not lexically North Bahnaric.

5 Conclusion

With the exception of the nasal hardening shared with Kơtua, it is apparent that Cua shows no distinctive phonological or lexical features that might connect it with other Bahnaric sub-branches. Taking all of the above into consideration, we must ascribe the structural parallel with Kơtua to contact, linguistic tendency, or some other non-genetic mechanism. Cua is best accounted for historically by reconstructing its direct descent from Proto Bahnaric, constituting its own Eastern sub-branch. This leads me to offer the Bahnaric classification in Table 9, essentially the same as Sidwell (2002), with Cua classified alone as East Bahnaric.

[11] This is, of course, in the absence of sources that might facilitate an adequate examination of the rather scant morphology.

Table 9: Revised Bahnaric classification by Sidwell (this paper)

West Bahnaric	Central Bahnaric	North Bahnaric	East Bahnaric
Jru' (Laven), Juk, Su'	Taliang (Kasseng)	Halang, Kayong	Cua (Kor)
Nyaheun	Alak	Jeh	
Oi, The, Sok, Sapuan, Cheng	Central South	Kotau	
Brao, Laveh, Krưng, Kravet	Tampuon	Tadrah, Modrah	
	Bahnar	Sedang	
	South Bahnaric	Hrê	
	Chrau	Mơnơm (Bơnâm)	
	Sre	Rengao	
	Stieng	Kaco', Ramam	
	Mnong		

By way of closing remarks, a significance aspect of the present result is that we may confirm that the geographical centre of Bahnaric diversity is more or less right where the North Bahnaric languages are located today, on the plateau around and north of Kontum. This is rather close to the centre of diversity of the Katuic languages, approximately a hundred kilometers further north-west, and adjacent to the coastline where we believe the ancient Chams first established themselves on the Indo-Chinese mainland.

References

Blench, Roger. 2008. Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology? In Alicia Sanchez-Mazas, Roger M. Blench, Malcolm Ross, I. Peiros and Marie Lin, eds *Human migrations in continental East Asia and Taiwan. Matching archaeology, linguistics and genetics*, 105–132. London: Routledge.

Blood. 1966. A reconstruction of Proto-Mnong (including tentative reconstruction of Proto-South-Bahnaric). MA thesis. Bloomington: Indiana University.

Burton, Eva. 1969. A brief sketch of Cua clause structure. *Mon-Khmer Studies* 3.5–8.

Cooper, James and Nancy Cooper. 1966. Halang phonemes. *Mon-Khmer Studies* 4.87–98.

Diffloth, Gérard and Norman Zide. 1992. Austro-Asiatic languages. In William Bright, ed. *International Encyclopædia of Linguistics,* 137–142. Oxford: Oxford University Press.

Gregerson, Kenneth J., Kenneth D. Smith and David D. Thomas. 1976. The place of Bahnar within Bahnaric. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta, eds *Austroasiatic studies, part I*. Oceanic Linguistics Special Publication, No.13, 371–406. Honolulu: University of Hawaii Press.

Lebar, Frank M., Gerald C. Hickey, and John K. Musgrave. 1964. *Ethnic groups of mainland Southeast Asia*: New Haven Connecticut: Human Relations Area Files Press.

Maier, Jacqueline. 1969. Cua phonemes. *Mon-Khmer Studies* 3.9–19.

Maier, Jacqueline and Eva Burton. 1971. Vietnam word list (revised): Cua. Unpublished manuscript. Bangkok: David Thomas Library (folder 495.9).

—— 1981. Cua–English–Vietnamese dictionary: 1981 pre-publication manuscript based on a 1966 manuscript. Unpublished manuscript. Dallas: Summer Institute of Linguistics.

Maier, Jacqueline and Đình Van Cau. 1976. Cua vocabulary: Cua-Vietnamese-English Tthesaurus (a computer print-out). Unpublished manuscript. Huntington Beach California: Summer Institute of Linguistics.

Phillips, Richard. 1959. Vietnam tribes survey word lists: Cua (=Traw). Unpublished manuscript. Bangkok: David Thomas Library (folder 495.968).

Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*: Wiesbaden: Otto Harrassowitz.

—— 1963. The position of the Munda languages within the Austroasiatic language family. In H.L. Shorto, ed. *Linguistic comparison in Southeast Asia and the Pacific*, 140–152. London: School of Oriental and African Studies.

Sidwell, Paul. 2002. Genetic classification of the Bahnaric languages: a comprehensive review. *Mon-Khmer Studies* 32.1–24.

—— n.d. Proto North Bahnaric: a reconstruction of phonology and lexicon. Unpublished manuscript. Canberra: The Australian National University.

Sidwell, Paul and Pascale Jacq. 2003. *A handbook of comparative Bahnaric:* vol. 1, *West Bahnaric*: Canberra: Pacific Linguistics.

Smith, Kenneth D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Language Data Series. Dallas: Summer Institute of Linguistics.

—— 1973. Eastern North Bahnaric: Cua and Kơtua. *Mon-Khmer Studies* 4.113–118.

Theraphan, Luang-Thongkum. 2001. *Languages of the tribes in Xekong Province Southern Laos*. Bangkok: The Thailand Reseach Fund.

Thomas, David. 1966. Mon-Khmer subgroupings in Vietnam. In Norman H. Zide, ed. *Studies in Comparative Austroasiatic Linguistics*, 194–202. The Hague: Mouton.

—— 1979. The place of Alak, Tampuon, and West Bahnaric. *Mon-Khmer studies* 8.171–186.

Thomas, David and Robert Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25.398–418.

11 *Binanderean as a member of the Trans New Guinea family*

JACINTA SMALLHORN

1 Introduction[1]

Greater Binanderean is a group of genetically related Papuan languages of southeast Papua New Guinea. In this paper I seek to demonstrate what lexical evidence there is for Greater Binanderean's inclusion in the putative Trans New Guinea (TNG) family, the largest posited family of non-Austronesian languages in the New Guinea area.

1.1 The Greater Binanderean group

The Binanderean group in its entirety was identified by Wilson (1969). It consists of the following languages, running in approximately this order from north to south: Suena, Yekora, Zia, Mawae, Binandere, Ambasi, Aeka, Orokaiva (dialects), Hunjara, Notu, Yega, Gaena, Baruga (dialects), Doghoro, and Korafe. The area covered by the languages extends from around the Mai-Ama River in the southeastern corner of Morobe Province down to Siu Village on Cape Nelson (where Korafe is spoken) in Oro (or Northern) Province.

Guhu-Samane (or Mid Waria), long recognised as a language closely related to the Binanderean group, is at the northern end of this area on the Pai-Awa River and to the west on the Mai-Ama River and beyond the Bowutu coastal mountain range. Using lexicostatistical evidence, Hooley and McElhanon (1970:1075–1076) classified this 'isolated' language as the closest relative of the Binandere Language Family, thereby creating the Binandere Stock. In the terminology of the present paper, *Greater Binanderean* (a term not previously used) equates to Hooley and McElhanon's Binandere Stock, whereas *Binanderean* refers to the family-level group (i.e. excluding Guhu-Samane).[2]

Conceptions of a linguistic group corresponding to the greater part of the Binanderean group as it is presently known go back more than a century (see, for example, Seligmann 1906). Cognate percentages given for Binanderean languages range from 52 to 86% for

[1] This paper is largely based on a chapter of Smallhorn (in prep). I am indebted to Andrew Pawley for revising this paper and to Edgar Suter for providing unpublished data and comments. I am also grateful to Bethwyn Evans and Bevan Barrett for their comments.

[2] In the lexicostatistical classification, 'stock' signified between 12% and 28% putative cognates and 'family' between 28% and 55%.

Suena, Yekora, Zia, Mawae, and Binandere (Wilson 1969:76) and from 50 to 60% for Notu, Gaena, Baruga, Doghoro, and Korafe (Dutton 1971:9–11). Capell (1962:147–148) likened the relationship between the Binanderean varieties of Northern Province (see Map 1) to that between the Romance languages of Europe.

Approximately 80,000 people speak languages of the Greater Binanderean group.[3] (See Map 1 for the locations of these languages.)

Map 1: The Greater Binanderean language group of southeast Papua New Guinea

[3] This figure is based on figures listed against the individual languages on the internet Summer Institute of Linguistics in papua New Guinea (www.pnglanguages.org). Most of the figures for the languages are from the 2000s, and the remaining few from the 1970s, 1980s and 1990s. (Note that some members, listed in this paper as individual languages, are conflated in the source with those of which they can be considered dialects.)

Language groups adjacent to Greater Binanderean languages are Goilalan, Yareban, and Koiarian, all of which form part of the Southeast Papuan family, a putative Trans New Guinea subgroup (see §1.2). In Map 1 it can be seen at which points Greater Binanderean shares a border with members from these three groups: Guhu-Samane borders Goilalan, Orokaiva and Hunjara border Koiarian, and Baruga borders Yareban. Where Greater Binanderean does not share a border with one of these groups, it is almost entirely bordered by uninhabited areas.

1.2 The Trans New Guinea family[4]

The term 'Trans New Guinea' was first applied by McElhanon and Voorhoeve (1970) to a proposed grouping of the Finisterre-Huon family and a number of groups of Central New Guinea. This connection was based on around 90 possible cognate sets (representative of 53 meanings) in basic vocabulary. On the basis of putative cognates with 30 of these sets, Binanderean was foreshadowed by the authors as a possible subgroup of the family.

In a later, much expanded version of the TNG hypothesis, based on research at The Australian National University between 1970 and 1975, Wurm, Voorhoeve and McElhanon, in a paper in Wurm's large volume (ed. 1975), posited a group composed of around 256 languages considered to be secure members. In it are all the languages of the central cordillera east of the Bird's Head, from the Wissel Lakes and the Baliem Valley to southeastern New Guinea, as well as the Finisterre-Huon group and some other languages spoken to the north of the central ranges, and a few spoken to the south (mainly the Asmat-Kamoro, Awyu-Dumut, and Lowland Ok groups). The groups of southeastern New Guinea (Binanderean and Guhu-Samane, Goilalan, Koiarian, Kwalean, Manubaran, Yareban, Mailuan, and Dagan) were grouped together as the Eastern Part of the Trans-New Guinea Phylum's 'Main Section'. The authors' diagnostics for inclusion were lexicostatistical comparisons, about ten putative cognate sets in very stable and widely distributed basic lexicon, three sets of free form pronouns, and select structural features in morphology and syntax.

Also in Wurm (ed. 1975) appeared Dutton's article on the southeastern languages of the Trans New Guinea family, i.e. Goilalan, Koiarian, Kwalean, Manubaran, Mailuan, Dagan, Yareban, and the Binanderean Stock. Herein Dutton outlined some of the general phonological, morphological, and syntactic characteristics of the languages abstracted from available sources. He then proceeded to a brief treatment of their lexical similarities, giving cognate percentages. He concluded that there was no clear-cut lexical evidence, based on his sample, for subgrouping the eight southeastern groups in any one particular way above the family level, adding, moreover, that typological patternings further confused the picture. This conclusion was an amendment to Dutton's (1969) earlier hypothesis, put forward in a paper by Wurm (1971:565–571), of a South-East New Guinea Phylum, tentatively incorporating Koiarian, Manubaran, and Yareban (these three proposed as forming a stock, along with the 'family-type language' Kwale), Goilalan, and Binanderean, with Mailu provisionally included and Daga foreshadowed as possibly figuring in the grouping with further study.[5]

---

[4] This subsection follows in large part Pawley (2005).

[5] Long before, Strong (1911) had pointed to a possible relationship of Binanderean with Mailuan and 'Upper Musa', which constituted languages now known as belonging variously to Yareban, Mailuan, and Baruga/Doghoro of the Binanderean group.

In the aforementioned article by Wurm, Voorhoeve, and McElhanon (1975), the authors posited a less conservative, more sweeping TNG group, which encompassed a total of 491 languages, almost double the number included in the more conservative group. Added were the following groups: Madang-Adelbert Range, Border, Eleman, Inland Gulf, Kalam, Kolopom, South Bird's Head, Teberan-Pawaian, and Trans Fly. The Papuan languages of Timor, Alor, and Pantar in eastern Indonesia were also added but assigned to the marginal category. Languages were assigned to the marginal category if their structure differed from typical TNG structure and if they showed a paucity of specific lexical resemblances with typical TNG languages. These languages were regarded by the three scholars as hybrids resulting from an overlaying of a TNG component on a non-TNG substratum.

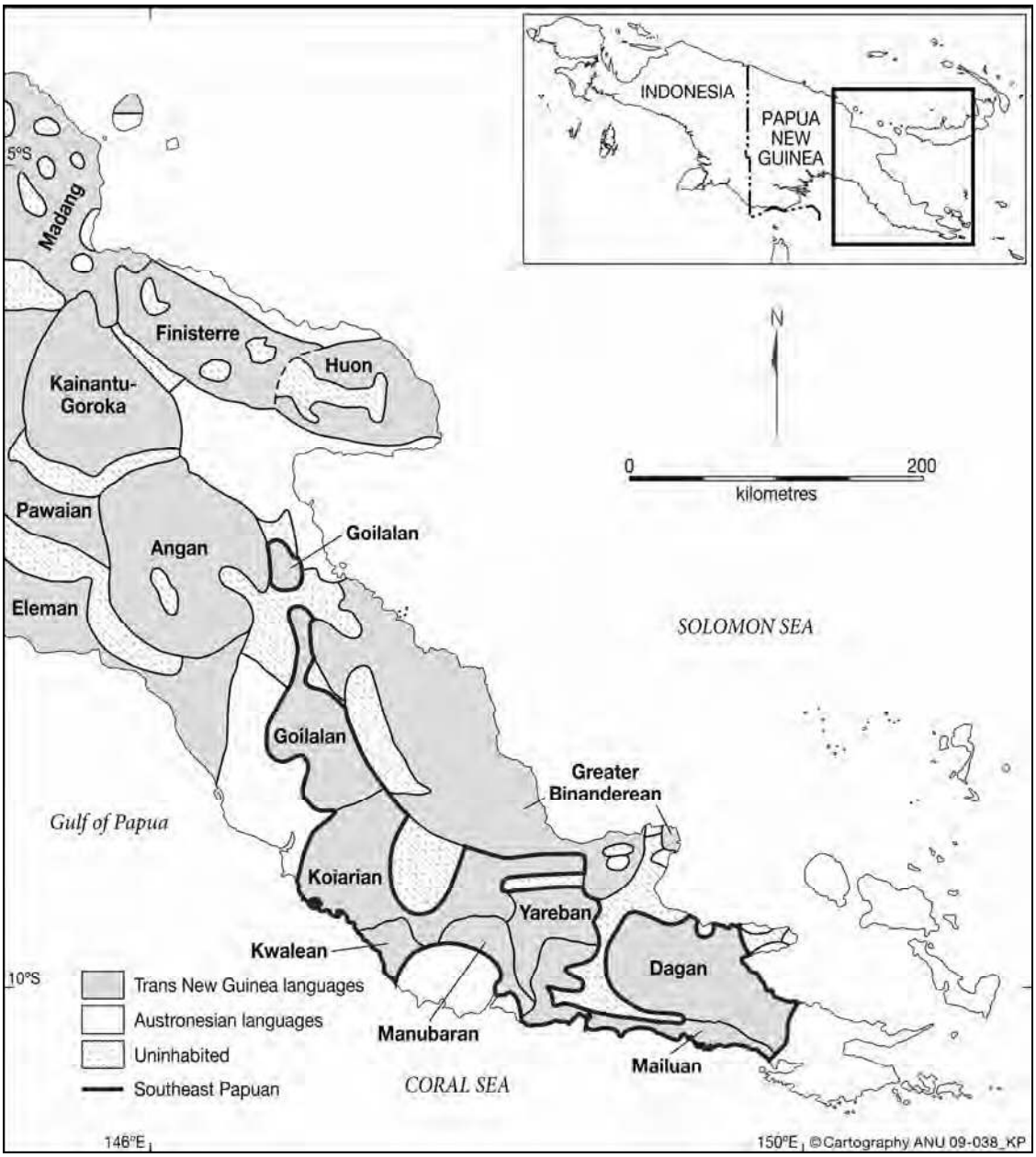
In the 1990s work was resumed on the TNG hypothesis, this time with more emphasis on use of the comparative method and employing both top-down and bottom-up approaches. The latest incarnation of the hypothesis is based on: (a) a body of (more or less) regular sound correspondences, which in turn are based on (b); (b) approximately 200 putative cognate sets of basic vocabulary occurring in two or more major subgroups; (c) systematic form-meaning correspondences in the personal pronouns, permitting reconstruction of virtually a complete paradigm; and (d) widespread resemblances in fragments of certain other grammatical paradigms (Pawley 2005:81, 92). In this latest version of the TNG family Greater Binanderean has been included on the basis of lexical evidence (Pawley 2000a, 2001) and fragmentary evidence in Binanderean from a subject suffix paradigm reconstructed for an early stage of TNG, ancestral to at least several disparate subgroups of North-East New Guinea, and Madang, Morobe, and Eastern Highlands Provinces (Suter 1997; Pawley 2000b, 2005:89–91); though not on the securest diagnostic evidence, on which work has been carried out by Ross (1995, 2005, in prep), namely, the personal pronoun paradigm. Ross reserves judgment on the TNG status of Greater Binanderean (Ross 2005:36–37, 51) since this subgroup appears to reflect only one of the TNG pronouns, namely, *na '1S'. The fact that the sole pronoun reflected is a monosyllable with a nasal consonant makes the resemblance even flimsier as evidence (see, for example, Campbell 1997:240–252). Ross has renamed the grouping of the seven other southeastern families Southeast Papuan, including it as a possible subgroup of TNG defined by the innovative pronominal *ya '2P'. He has excluded Greater Binanderean, which does not participate in this innovation (Ross 2005:23; in prep).

1.3 Methodology

In order to demonstrate Greater Binanderean's likely membership in the putative TNG family, I will first present Proto Trans New Guinea (pTNG) etyma that have been retained in the group. Where these etyma appear to be retained also in the TNG subgroups Huon,[6] Madang and Koiarian (see Map 2), reflexes from selected languages in these three groups will also be shown. Some possible cognate comparisons of Greater Binanderean forms not proposed as TNG with forms from these three other subgroups will then be presented, constituting further possible evidence of Greater Binanderean's membership in the wider, TNG family. While these sets may constitute forms which could be uniquely shared by Greater Binanderean and that subgroup, closer inspection of other groups, though not undertaken here, may yield resemblant forms there as well.

---

[6] The Huon languages only, rather than the combined Finisterre-Huon group, have been chosen as I have been able to draw on comparative work on this group by Edgar Suter.



Map 2: Trans New Guinea language groups of southeast New Guinea

In a very small number of cases a Guhu-Samane form will be given where no cognate form is known to occur in Binanderean languages. For many of the Binanderean reconstructed forms that are given, cognate Guhu-Samane forms occur but will not be shown.

Goilalan and Yareban, the two Southeast Papuan groups besides Koiarian which border parts of Greater Binanderean territory (see Map 2), are not as well studied as the aforementioned TNG subgroups. Due to considerations of space a comparison of these lesser-known groups with Greater Binanderean will not be attempted here.

Two other putative TNG subgroups, Angan and Kainantu-Gorokan, are also in the vicinity of the Greater Binanderean languages, lying to the northwest. Angan is not considered to be closely related to any other TNG group, having a 5% lexical resemblance to Kainantu-Gorokan to its north and 3% to Pawaian to its west, nor is the Eleman group to

its south (Lloyd 1973:33; Osmond et al. in prep). The Kainantu-Gorokan languages, originally placed in an East New Guinea Highlands stock during the early development of the TNG hypothesis, are not currently considered to be more closely related to any one TNG group over any other.

The Binanderean reconstructions given in this paper are mostly at the level of either Proto Binandere (pBin) or Proto Nuclear Binandere (pNucBin), an interstage from which the majority of Binanderean languages stem. In a few cases the form given is that from a lower-order subgroup or present-day member language. The following internal grouping of the Greater Binanderean group is proposed (with some lower-level detail omitted), with abbreviations given for the stages used thereafter.[7]

PROTO GREATER BINANDERE (pGBin)
- Guhu-Samane
- PROTO BINANDERE (pBin)
  - *Proto North Binandere* (pNBin): Suena, Zia, Mawae
  - Yekora
  - *Proto Nuclear Binandere* (pNucBin)
    - *Proto Binandere-Ambasi*: Binandere, Ambasi
    - *Proto South Binandere* (pSBin)
      - *Proto Orokaiva*: Aeka, Orokaiva, Hunjara
      - *Proto Coastal Binandere* (pCstBin): Notu, Yega, Gaena, Baruga, Doghoro, Korafe

Figure 1: Internal grouping of Greater Binanderean

2 Retentions of TNG etyma in Greater Binanderean that are shared with other TNG subgroups

Based on evidence pointed to above as well as on the findings of the present paper, I conclude that the Greater Binanderean languages belong to the TNG family. Proto Trans New Guinea reconstructions with their putative reflexes in Binanderean (or Guhu-Samane) are set out in Table 1 (ordered alphabetically by the pTNG gloss), with a small selection of corresponding forms from Huon, Koiarian, and Madang where these have been found to occur. Pawley lists a total of 190 pTNG forms (2005:85–88); only those with potential cognates in Greater Binanderean are included in the table. The Huon, Koiarian, and Madang forms marked (P) are from Pawley (2001). The Huon forms marked (S) are from Suter (2007). The Koiari forms marked (D) are from Dutton (2003). The Madang forms marked (Z) are from Z'graggen (1980a,b,c,d[8]), with abbreviations for language groups/ sources as follows: NAR: Northern Adelbert Range; SAR: Southern Adelbert Range; Mab: Mabuso; Rai: Rai Coast. The pronouns are taken from Ross (in prep).

[7] Reconstructed forms in Binanderean and internal subgrouping are taken from Smallhorn (in prep).

[8] a: Rai Coast; b: Northern Adelbert Range; c: Mabuso; d: Southern Adelbert Range.

Table 1: Putative reflexes of pTNG etyma in Greater Binanderean, with corresponding Huon, Koiarian and Madang forms

Proto Trans New Guinea	Binanderean/ Guhu-Samane	Huon	Koiarian	Madang
*mundun 'belly, internal organs'	pBin *{m,mw}unju 'egg'	Kâte *munduŋ* 'inner yolk' (P)		Dumpu (Rai) *munu* 'heart' (P)
*pu + verb 'blow'	pBin *pus- 'blow'	Lower Mape *pute* 'blow' (S)		Biyom (SAR) *fu-* 'blow on (fire)' (Z),[9] Ulingan (NAR) *fur-* 'blow on (fire)' (Z)
*kondaC 'bone'	Binandere *undoru* 'bones, skeleton'[10]			Waskia (NAR) *kutiŋ* 'bone' (P)
*amu 'breast'	pBin *ami 'breast'	Ono *ŋɑmu* 'breast, milk' (S), Selepet *nam* 'milk, breast, teat' (S)	Koiari *ɑmu* 'breast; milk' (D)	Suroi (Rai) *amo* 'breast (woman)' (Z)
*apa[pa]ta 'butterfly'	Guhu-Samane *qipopo* 'butterfly, moth'[11]			Pila (NAR) *kapopɔr* 'butterfly' (Z), Asas (Rai) *papʌu* 'butterfly' (Z)
*kumV- 'die'	pBin *ambu- 'wither, be sick, be dying'	Ono *koma* 'die (of fire)' (S), Burum *kəmu* 'die' (S)		Suroi (Rai) *kumi-* 'die' (Z)
*ka(nd,t)(i,e)C 'ear'	Yega *kari* 'ear' (P)	Kâte *hatseʔ-* 'ear' (P), Burum *ketsap* 'ear' (S)	Koita *korema* 'ear' (P)	Sileibi (SAR) *kinziŋa* 'ear' (P)
*na- 'eat, drink'	Suena *nai-* 'eat'[12]	Ono *ne* 'eat' (S), Burum *ne* 'eat, drink' (S)		Sinsauru (Rai) *n-* 'eat' (Z)
*(ŋg,k)iti [maŋgV] 'eye'	pBin *giti 'head' *mokə 'centre'			Kalam *[wdn] magi* 'eye' (P)
*apa 'father'	pSBin *apa 'father'			Kalam *-ap* 'father' (P)
*mbapa 'father'	pNucBin *baba 'father'	Nabak *bip* 'father' (P)	Koiari *baba* 'classificatory father, dad (as form of address)' (D)	Kalam *bapi* 'father' (P)
*mbalaŋ 'flame'	pBin *beriberi 'flame'	Ono *bolam* 'flame' (S), Selepet *balam* 'flame, torch' (S)		Kalam *malaŋ* 'flame' (P)

9 Biyom is revised in Pawley (2001) as grouping with the Rai Coast languages.
10 King (1927:51).
11 Richert and Hoopusu (2002:176).
12 Farr and Larsen (1979:10).

Proto Trans New Guinea	Binanderean/ Guhu-Samane	Huon	Koiarian	Madang
*mV- ’give’	pBin *mut- ‘put, give’		Koiari *mo* ‘give’ (D)	Garus (Mab) *ma-* ‘give’ (Z), Munit (Mab) *me-* ‘give’ (Z)
*yata- ‘go’	pCstBin *y- ‘go’			Sausi (Rai) *yarɛ-* ‘go’ (Z), Kesawai (Rai) *yan-* ‘go’ (Z)
*sikal/*sakil ‘hand, claw’	pBin *siŋgo ‘finger’ (Guhu-Samane *sika* ‘little finger’[13])			Mawak (NAR) *siŋgɛn* ‘fingernail’ (Z)
*kV(mb,p)utu ‘head’	pNucBin *kopiru ‘head’	Kâte *kpitseʔ* ‘head’ (P)		Suroi (Rai) *gabat* ‘head’ (P)
*ambi ‘husband, man’	pBin *embə ‘man’	Burum *ambi* ‘woman’ (P), Selepet *ibi* ‘woman’ (P)	Managalasi *ema* ‘married man’ (P)	Kalam *b* [mbə] ‘man’ (P)
*na ‘I (1S pronoun)’	pBin *na ‘I (1S pronoun)’	pHuon *na ‘I (1S pronoun)’	pKoiarian *na/*d[a,i] ‘I (1S pronoun)’	
*nVŋg- ‘know, hear, see’	pBin *niŋg- ‘hear’	Selepet *naga~nagɑ~nagan* ‘listen, hear, know, feel, understand’ (S)		Kalam *nŋ-* ‘see, hear, know’ (P), Garuh (Mab) *neg-* ‘watch’ (P)
*(mb,m)elak ‘lightning, light’	pBin *birigi, *{p,b}iri{p,b}iri ‘lightning’	Kâte *babariʔ* ‘lightning, brightness’, *bariʔ* ‘sparkle, glitter, glory’ (S), Selepet *belek* ‘lightning’ (S)		Ukuriguma (NAR) *bilika* ‘lightning, lightning flash’ (P), Bau (Mab) *peri [flay-]* ‘flash, lighten’ (P)
*k(o,u)t(u,i)p ‘long’	Mawae *gorebu* ‘long’[14]	Burum *kərip-ŋi*, *kərəp-ŋi* ‘long, high, far’ (S), Selepet *kalip* ‘long’ (S)		Urigina (Rai) *uruβa* ‘long’ (P)
*muk ‘milk, sap’	pBin *mu ‘sap’	Kâte *moŋ* ‘milk, breast’ (P)		Kalam *muk*, *mok* ‘milk’ (P)
*at(i,u) ‘netbag’	pBin *ati ‘string bag’			
*k(i,u)tuma ‘night’	pBin *tumba ‘night’			Rapring (Mab) *tumag* ‘night’ (P)
*mundu ‘nose’	pBin *mwendə ‘nose’			Munit, Sihan (Mab) *mede* ‘nose’ (P)

[13] Richert and Hoopusu (2002:201).
[14] McElhanon and Voorhoeve (1970:89).

Proto Trans New Guinea	Binanderean/ Guhu-Samane	Huon	Koiarian	Madang
*mbeŋga-[masi] 'orphan, widow'	Suena *boga masa* 'destitute'[15]	Kâte *bekɔ* 'orphan', *mɔsiŋ* 'widow and orphans' (P)		Amele (Mab) *beg-a-beg* 'widow and orphan(s)' (P)
*[mb]amba 'sibling, older same sex'	pNucBin *ambo 'younger sibling or lastborn'			Kolom (Rai) *aba* 'sibling, same sex' (P)
*aya 'sister'	pBin *ai, *aŋa 'mother, mother's sister'			Kolom (Rai) *ana* 'mother' (Z)
*kambu(s,t)(a,u) 'smoke'	pBin *imbosi 'smoke'	Ono *kopu* 'smoke' (P)		Biyom (Rai) *butu* 'smoke' (P)
*kasipa- 'spit (v.)'	pNucBin *kosiwa 'spittle'	Burum *səwət* 'spittle, saliva' (S)		Tauya (Rai) *sipina* 'saliva' (P)
*[na]muna 'stone'	pNucBin *ganuma 'stone'			Ulingan (NAR) *nomona* 'stone' (Z)
*kamb(a,u)na 'stone'	pBin *gomba(ro)/ *gembi(ro) 'stone'	Kâte *kpana* 'stone' (P)		Urigina (Rai) *goimbo* 'stone' (P)
*asi 'string, rope'	pBin *asi(N) 'vine, rope, string'			Atemple (SAR) *asʉn* 'rope (of bow)' (Z)
*mV 'taro'	pBin *mba 'taro'			Usino, Sumau (Rai) *mɛ* 'taro' (Z), Usu (Rai) *ma* 'taro' (Z)
*ndaŋgi-/ *ndiŋga- 'tie'	pBin *di- 'tie'	Burum *dzəhə* 'bind, tie' (S)		Samosa (Mab) *diga-* 'tie' (P)
*titi 'tooth'	pBin *di 'tooth'	Burum *dzit*, *dzət* 'tooth, mouth' (S)		Mugil (NAR) *-te* 'tooth' (Z)
*mbuli[ki] + verb 'turn (oneself)'	Guhu-Samane *burisi eetaqu* 'to turn over or to turn it around'[16]			
*[si]si 'urine'	pBin *pusisi 'bladder', *susu 'urine'	Kâte *fazi* 'urine', *fafazi* 'bladder' (S)		Girawa (Mab) *si* 'urine' (Z)
*panV 'woman, female'	pBin *bam{u,o}nə 'woman'			Kalam *pañ*, *pay* 'girl' (P)

[15] Wilson (1980:14).
[16] Richert and Hoopusu (2002:386).

This sizeable number of retentions of pTNG etyma in Binanderean (or Guhu-Samane), i.e. 39 forms in Table 1 (a fifth of Pawley's proposed list of TNG etyma), almost all representing basic vocabulary, and the consequent proposal of sound correspondences, given in Table 2, constitute compelling evidence for treating the group as a branch of TNG. Note that although the Binanderean column of Table 2 is headed 'pBin', a number of the Binanderean items to which the correspondences refer are reconstructed to a lower level than Proto Binandere or are from a present-day language. Among these, two items (Suena *nai-* 'eat', pNucBin *baba 'father') are included in the sound correspondences based on the external evidence alone, as the pBin consonants cannot be reconstructed from the internal evidence; and two items (Suena *boga masa* 'destitute', Guhu-Samane *burisi eetaqu* 'to turn over or to turn it around') are excluded from the sound correspondences since neither the internal nor the external evidence can be used to ascertain what the pBin consonants would have been.

The variability occurring in many of the pBin correspondences, i.e. in those corresponding to pTNG *ŋg, *m, and the voiceless plosives, cannot be accounted for by conditioning on the basis of so few resemblant pairs. This degree of variability is common among language groups exhibiting putative reflexes of pTNG (Pawley 2001:295).

Table 2: Tentative consonant correspondences between pTNG and Binanderean

pTNG	pBin	Instances (only pTNG forms shown)
*p	*p	***p**u, *a**p**a[**p**a]ta, *a**p**a, *kV(mb,**p**)utu
	*b	*mba**p**a, *k(o,u)t(u,i)**p**, ***p**anV
	*w	*kasi**p**a-
*t	*t	*(ŋg,k)i**t**i [maŋgV], *k(i,u)**t**uma, *a**t**(i,u)
	*d	***t**iti
	*r	*ka(nd,**t**)(i,e)C, *kV(mb,p)u**t**u, *k(o,u)**t**(u,i)p
*k	*k	***k**a(nd,t)(i,e)C, ***k**V(mb,p)utu, *si**k**al/*sa**k**il, ***k**asipa-
	*g	*(ŋg,**k**)iti [maŋgV], *(mb,m)ela**k**, ***k**(o,u)t(u,i)p, ***k**amb(a,u)na
*k-	*Ø-	***k**ondaC, ***k**umV-, ***k**ambu(s,t)(a,u)
*mb	*b-	***mb**apa, ***mb**alaŋ, *(**mb**,m)elak
	*-mb-	*a**mb**i, *[mb]a**mb**a, *ka**mb**u(s,t)(a,u), *ka**mb**(a,u)na
*nd	*d-	***nd**aŋgi-/***nd**iŋga-
	*-nd-	*ko**nd**aC, *mu**nd**u
*ŋg	*k	*(ŋg,k)iti [ma**ŋg**V]
	*ŋg	*nV**ŋg**-
*s	*s	****s**ikal/***s**akil, *ka**s**ipa-, *kambu(**s**,t)(a,u), *a**s**i, *[**s**i]**s**i
*m	*m	****m**undun, *a**m**u, *(ŋg,k)iti [**m**aŋgV], ***m**V-, ***m**uk, *[na]**m**una
	*mb	*ku**m**V-, *k(i,u)tu**m**a, ***m**V
*n	*n	****n**a-, ***n**a, ***n**Vŋg-, *[na]mu**n**a, *pa**n**V
*l	*r	*mba**l**aŋ, *(mb,m)e**l**ak
*y	*y	****y**ata-, *a**y**a (pBin: -i)
*-C	*-Ø	*mundu**n**, *mbala**ŋ**, *sika**l**/*saki**l**, *mu**k**

Table 3 displays potential cognate forms between Binanderean and two of our three other representative TNG subgroups. These are not reflexes of proposed pTNG etyma but may form the basis for further pTNG reconstructions due to their multiple occurrence in putative subgroups. (The starred forms in the Koiarian column are taken from Dutton (in press); Proto Koiaric is a descendant of Proto Koiarian in Dutton's subgrouping of the Koiarian family.)

Table 3: Further forms possibly reconstructable to pTNG

Binanderean	Huon	Koiarian	Madang
pBin *tatau 'sibling/ father's sister'	Burum *dat* 'older sibling of same sex' (S)	Proto Koiaric *tata 'sibling, opposite sex, older'	
pNucBin *tomboru 'cassowary'		Proto Koiarian *tubu[are] 'cassowary'	Sinsauru (Rai) *tomoro* 'cassowary' (Z)
pBin *gosə 'cloud'	Burum *kousu* 'cloud, fog' (S)	Koiari *gousa* 'fine misty rain, fog'	
pBin *ce 'excreta'[17]	Selepet *tep* 'stomach, belly' (S)	Koiari *de* 'faeces, manure, excreta, dung' (D)	
pBin *burə 'garden, work'	Burum *bərə* 'work' (S)	Proto Koiarian *buru 'garden'	
pNucBin *anda 'lefthand'	Ono *kana* 'left' (S), Selepet *kane* 'left[handed]' (S)	Proto Koiarian *ada 'arm, hand, branch, wing, handle, foreleg (of animal)'	
pNucBin *mwandə 'women's house'	Ono *mat* 'house, village, woman's house' (S), Selepet *emet* 'house, home; place' (S)		Kolom (Rai) *wʌnde* 'house' (Z), Amaimon (NAR) *mɔnɛ* 'house' (Z)

3 Some resemblant forms shared by Greater Binanderean and Huon which lack proposed pTNG antecedents

The Huon languages, situated around the Huon Peninsula, lie north of the Greater Binanderean languages and are separated from them by a large area, partly uninhabited and partly occupied by speakers of Austronesian languages. They and the Finisterre languages form the Finisterre-Huon family. In Table 4 are some pairs of Binanderean (or Guhu-Samane) and Huon resemblant forms, the latter taken from Suter (2007), that lack proposed pTNG etyma. The representative Huon languages used for comparison are the Eastern Huon languages Ono and Kâte and the Western Huon languages Burum and Selepet, as these four have borrowed relatively little from other Huon languages, and as the latter two are phonologically the most conservative languages (Suter pers. comm.).

[17] The three forms in this set may actually be derived from pTNG **simbi* 'guts, intestines, bowels', from which, for example, Kalam *sb* [simp] 'excrement' (P) derives.

Table 4: Putative cognate sets that may be unique to (Greater) Binanderean and Huon

Binanderean/Guhu-Samane	Huon	Language
pBin *ma(n)i 'boy, son'	*medep* 'boy, child' *morə* 'offspring, young animal; very small'	Ono Burum
pBin *wa(N) 'canoe'	*woke* 'boat, canoe' *waŋge* 'boat, canoe' *waga* 'canoe, ship, wooden trough'	Kâte Burum Selepet
pNucBin *pur- 'come'	*fa~fa* 'come (to you)'	Kâte
pBin *diti, *dibe 'eye'	*dzaŋe* 'eye' *dze* 'eye'	Kâte Burum
Guhu-Samane *boto* 'hand, arm'[18]	*bat* 'arm, hand'	Selepet
pBin *mbedi 'hungry'	*bodi* 'famine'	Ono
pBin *soro 'index finger'	*halip* 'fingers, small toes'	Selepet
pBin *paC- 'lie'	*fo* 'lie, be there'	Kâte
pNucBin *siŋga 'male youth'	*zeŋoŋ* 'unmarried youth, virgin' *sihan* 'young, unmarried'	Kâte Selepet
pBin *g- 'see'	*ka* 'see it'	Ono
pBin *utumu 'shade'	*umut* 'shadow, image, picture'	Selepet
pBin *me 'shame'	*mimi* 'shame, sense of shame'	Kâte
pBin *waeko 'sun'	*wehən* 'sun, day'	Burum

Note that the Binanderean and Huon terms for 'canoe' have been included in the comparison although they are clearly of foreign origin: Proto Oceanic, the ancestor of the Austronesian languages of New Guinea, gives us *waga 'canoe' (Ross et al. 1998:178). Suter (pers. comm.) suggests that 'canoe' in Huon is virtually the only Oceanic loanword possibly reconstructable for Proto Huon Peninsula, and it does seem to be possibly reconstructable for Proto Binandere as well. The inclusion of the forms denoting 'canoe' in the Binanderean-Huon comparison would only be valid if they were reflexes of an ancestor language which was common to both Binanderean and Huon, but such an ancestor would very likely predate the arrival of Oceanic speakers on the New Guinea mainland; and if this were the case, we would expect far more resemblances between Binanderean and Huon than we actually find. Thus, the occurrence of a resemblant form in the two language groups is likely to be a case of independent borrowing.

4 Some resemblant forms shared by Binanderean and Koiarian which lack proposed pTNG antecedents

As earlier mentioned, one of the three Southeast Papuan language groups bordering the Greater Binanderean group is Koiarian, for which detailed comparative work has been carried out by Dutton (1969, in press). It appears that Koiari, the best-described member of the Koiarian family, is phonologically conservative and therefore a good choice for comparison with Binanderean reconstructions. Table 5 presents some Binanderean and

[18] Richert and Hoopusu (2002:50).

Koiari resemblant forms, the latter taken from Dutton's (2003) dictionary of Koiari. In cases where a Koiarian reconstruction from Dutton (in press) is available, this is also given. (Proto Baraic is a descendant of Proto Koiarian in Dutton's subgrouping of the Koiarian family.)

Table 5: Putative cognate sets that may be unique to Binanderean and Koiarian

Binanderean	Koiari	Earlier Koiarian
Orokaiva dialects *sisae* 'bad'[19]		Proto Baraic *sise- 'bad'
Baruga dialects *tatangu, tatango* 'blood'[20]	*tavo* 'blood'	Proto Koiarian *taɣo 'blood'
pBin *ya(wa) 'dance'		Proto Koiarian *yava- 'dance, sing'
Baruga dialects *fouma, pouma* 'hair'[21]	*homo* 'hair, fur; feather'	Proto Koiarian *fómo 'hair, fur, feather'
pBin *d- 'hit'	*dai* 'bump, hit'	
pBin *ewa 'sea'	*eve* 'sea, salt water, salt'	
pNucBin *paro 'smoking platform'	*varo* 'stage for feast, sacred platform'	
Suena *omata* 'stone'[22]	*amata* 'soft flat stone, mud stone, shale'	

Note that the pairs with Binanderean meanings 'dance' and 'smoking platform' are problematic as they are cultural terms and are therefore susceptible to borrowing, and that 'bad' is not usually a stable term.

5 Some resemblant forms shared by Binanderean and Madang which lack proposed pTNG antecedents

The final comparison will be with various languages of the Madang group, which lies to the northwest of Greater Binanderean, and forms for which are taken from Z'graggen (1980a,b,c,d). This is exemplified in Table 6. The number and term given in the rightmost column correspond to Z'graggen's numbering and naming of lexical sets. Note that other possible resemblant pairs are apparent, but most of them look weaker and thus are not shown here; however, closer inspection may yield a greater number of viable pairs.

19 Farr and Larsen (1979:8).
20 Farr and Farr (1989).
21 Farr and Farr (1989).
22 Wilson (1980:44).

Table 6: Putative cognate sets that may be unique to Binanderean and Madang

Binanderean	Madang	Madang language/area	Madang reference in Z'graggen (1980)
pBin *bor{e,u} 'bamboo'	*borɛa*	Erima (Rai)	158: bamboo
	bʌ:ra	Kare (Mab)	
pNucBin *kambiya 'basket'	*kʌ:mo*	Duduela (Rai)	214: basket
	gumbʌ	Bom (Rai)	
	gama	Munit (Mab)	
pBin *urə 'belly'	*ouru*	Kesawai (Rai)	72: belly
	uřĩř	Malas (NAR)	
	uro	Wanuma (NAR)	
pBin *wugamba 'crocodile'	*waŋgaim*	Pulabu (Rai)	118: crocodile
	wʌŋgay	Yabong (Rai)	
	waga	Panim (Mab)	
pBin *tatamu/*tatauŋ 'drum'	*tʌmar*	Sileibi, Katiati (SAR)	206: hand drum
	tmauno	Biyom (SAR)	
	tʌmʌl	Amaimon (NAR)	
pBin *umbugə 'elbow'	*umbuŋgiŋguŋ*	Pondoma (SAR)	57: elbow
	uma	Ulingan (NAR)	
pBin *a 'female sex organ'	*a*	Bongu (Rai)	95: vulva
	ʌ:	Male (Rai)	
pBin *giti 'head'	*gate*	Bom (Rai)	33: head
	gati	Male (Rai)	
	gati-	Isebe (Mab)	
pNucBin *wewera 'hot'	*wɛra*	Kesawai (Rai)	236: hot
pNucBin *gomo 'liver'	*gʌma*	Sinsauru (Rai)	81: liver
	kɛma-	Sihan (Mab)	
	gʌm-	Murupi (Mab)	
	-gom	Korak (NAR)	
pBin *gor{i,o}bu 'marsupial mammal species'	*gob*	Saep (Rai)	121: rat
	gorʌwu	Parawen (NAR)	
pBin *tumə 'nape'	*tumʌku*	Sinsauru (Rai)	51: nape
	dumʌgu-	Asas (Rai)	
pBin *piŋgi 'pepper'	*pi:ni*	Kwato (Rai)	129: betel pepper vine
pBin *wa 'rain'	*wʌ*	Male (Rai)	191: rain
	wʌ	Amele (Mab)	
	va	Utu (Mab)	
pBin *utu 'sky'	*ʌtu*	Gumalu (Mab)	180: sky
	hutu	Sinsauru (Rai)	
pBin *kunə 'star'	*kunebi*	Asas (Rai)	185: star
	kɛnɛm	Usino (Rai)	
pBin *ndawa 'stick'	*nʌwa*	Sumau (Rai)	157: stick
pNBin *wari 'sun'	*wɛrɛ*	Sausi (Rai)	186: sun
	vir	Rapting (Mab)	
	woř̃ɛm	Waskia (NAR)	

Binanderean	Madang	Madang language/area	Madang reference in Z'graggen (1980)
	wɛdɛm	Saki (NAR)	
pNucBin *temb- 'walk'	*tɛmu-*	Tauya (SAR)[23]	310: walk
pBin *bubu(ra) 'wind'	*bubre* *oboř̃ɛ*	Suroi (Rai) Erima (Rai)	194: wind

Among these 'basket', being a cultural term, is not strong as a cognate set.

6 Conclusion

The earlier evidence for including Greater Binanderean in the Trans New Guinea family, as pointed out in §1.2, consists of possible cognates for about a third of McElhanon and Voorhoeve's (1970) 90 or so TNG etyma in basic vocabulary, later revisited by Pawley (2001), who included Binanderean or Guhu-Samane in about a quarter of his nearly 100 cognate sets; and of slight evidence in Binanderean from a reconstructed subject suffix paradigm. The present paper contributes to the question of Greater Binanderean's membership in TNG in giving: (i) an increased number of possible Binanderean (or Guhu-Samane) lexical cognates with TNG etyma, a large portion of which are reconstructed to the level of Proto Binandere; (ii) lexical comparisons with three other eastern New Guinea language groups which are putative members of TNG; and (iii) the suggestion that Greater Binanderean may be genetically closer to noncontiguous groups to its northwest than to neighbouring southeastern groups, with which it has been allied in some previous research. I additionally give further potential cognate sets representative of Binanderean, Huon, Koiarian, and Madang which may point to further TNG reconstructions.

An object of further research would be the search for innovations in phonology and morphology defining (Greater) Binanderean as a subgroup within TNG. One possible phonological innovation may be pBin *mb from pTNG *m. There are three instances which attest this (see Table 2). In the one case of these three where a Guhu-Samane cognate form is apparent (i.e. *maa* 'domesticated taro'[24] corresponding to pBin *mba 'taro', from pTNG *mV 'taro'), a simple nasal rather than a prenasalised stop is present. Two possibilities arise from this: (i) the ancestral consonant in pGBin was *m, inherited from pTNG *m, and the shift to *mb is therefore a pBin-defining innovation in which Guhu-Samane did not participate; (ii) the ancestral consonant in pGBin was *mb, but Guhu-Samane underwent a subsequent change back to *m* (a change which has occurred in pNBin [Smallhorn in prep]).

The lexical data in the present paper show Greater Binanderean as a likely member of TNG. From the evidence examined here, no conclusion can be reached as to which, of the Huon, Koiarian, and Madang groups, it may be more closely related to. Shared lexicon which constitutes retentions of TNG etyma is not evidence for subgrouping, but it is perhaps worthy of comment that based on the data herein presented, Greater Binanderean appears to share significantly more retentions with the groups to its north and northwest, Huon (28) and Madang (38), than it does with Koiarian (12) to its south and west (though this may be due to the large amount of organised data that exists for the Madang family).

[23] Tauya is revised in Pawley (2001) as grouping with the Rai Coast languages.
[24] Richert and Hoopusu (2002:120).

As to the number of items that are possibly unique to Greater Binanderean and another of these groups, the order is the same: Madang (20), Huon (13), Koiarian (8). Furthermore, comparing Greater Binanderean's comparisons with Madang and Huon, we find more resemblant phonemes per comparison in the Madang forms.

A closer genetic link with groups to the north would make sense when the likely Greater Binanderean homeland is taken into consideration. According to theories of maximum diversity and minimal population movements used to establish the most likely original dispersal centre (see, for example, Campbell 2004:400), the area of the boundary between the first-order subgroups, Guhu-Samane and Binanderean, is the most likely dispersal centre of the group.

In conclusion, it is well to point out that no-one has yet attempted an exhaustive comparison of the available lexical data for Trans New Guinea, and the lexical data in itself is patchy on the whole. There is also a need for more, local studies which will systematise the lexical data in the obvious, smaller groups of TNG (Lang 1976:74, Haiman 1979:897; Foley 1986:13) and provide a basis for pTNG lexical reconstruction. The caveat should also be offered that if the particular current hypothesis about the initial dispersal of the TNG family is correct — that this was associated with the spread of taro and banana cultivation (Denham 2005; Pawley 2005:97–100) between 4000 and 8000 BC — we are left with a time-depth of from six to ten thousand years. The earlier part of this range takes us back beyond the time-depth at which most comparativists think it is possible to reconstruct a protolanguage in detail.

References

Campbell, Lyle. 1997. *American Indian languages: the historical linguistics of native America*. New York: Oxford University Press.

—— 2004. *Historical linguistics: an introduction* (2nd edition). Edinburgh: Edinburgh University Press.

Capell, Arthur. 1962. *A linguistic survey of the south-western Pacific*. (New and revised edition.) Noumea: South Pacific Commission.

Denham, Tim. 2005. Agricultural origins and the emergence of rectilinear ditch networks in the highlands of New Guinea. In Pawley et al., eds, 2005, 329–361.

Dutton, T.E. 1969. *The peopling of central Papua: some preliminary observations*. Canberra: Pacific Linguistics.

—— 1971. Languages of south-east Papua: a preliminary report. *Papers in New Guinea linguistics, No. 14*, 1–46. Canberra: Pacific Linguistics.

—— 1975. South-eastern Trans-New Guinea Phylum languages. In Wurm, ed., 1975, 613–664.

— 2003. *A dictionary of Koiari, Papua New Guinea, with grammar notes*. Canberra: Pacific Linguistics.

— In press. *The Koiarian familiy revisited*. Canberra: Pacific Linguistics.

Farr, Cynthia and James Farr. 1989. Baruga vocabulary comparative list. Electronic manuscript. Ukarumpa: Summer Institute of Linguistics.

Farr, James and Robert Larsen. 1979. A selective word list in ten different Binandere languages. Unpublished manuscript. Ukarumpa: Summer Institute of Linguistics.

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

Haiman, John. 1979. Review of Wurm, ed. 1975. *Language* 55(4).894–903.

Hooley B.A. and K.A. McElhanon. 1970. Languages of the Morobe District – New Guinea. In S.A. Wurm and D.C. Laycock, eds *Pacific linguistic studies in honour of Arthur Capell,* 1065–1094. Canberra: Pacific Linguistics.

King, Copland. 1927. *Grammar and dictionary of the Binandere language*. Sydney: D.S. Ford.

Lang, Ranier. 1976. Review of Wurm, ed. 1975, *Kivung, Journal of the Linguistic Society of Papua New Guinea* 9(1).72–80.

Lloyd, Richard G. 1973. The Angan language family. In Karl Franklin, ed. *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*, 31–110. Canberra: Pacific Linguistics.

McElhanon, K.A. and C.L. Voorhoeve. 1970. *The Trans-New Guinea phylum: explorations in deep-level genetic relationships*. Canberra: Pacific Linguistics.

Osmond, Meredith and Andrew Pawley, Malcolm Ross. In prep. Membership and subgrouping of the Trans New Guinea phylum. In Pawley et al., eds, in prep.

Pawley, Andrew. 2000a. Some Trans New Guinea phylum cognate sets. Electronic manuscript. Canberra: The Australian National University.

—— 2000b. Notes on early Trans New Guinea verb morphology: reconstructing subject suffixes. Electronic manuscript. Canberra: The Australian National University.

—— 2001. The Proto Trans New Guinea obstruents: arguments from top-down reconstruction. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *Studies in Melanesian linguistics: papers in honour of Tom Dutton*, 261–300. Canberra: Pacific Linguistics.

—— 2005. The chequered career of the Trans New Guinea hypothesis: recent research and its implications. In Pawley et al., eds, 2005:67–107.

Pawley, Andrew, Robert Attenborough, Jack Golson and Robin Hide, eds. 2005. *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics.

Pawley, Andrew, Malcolm Ross and Meredith Osmond, eds. In prep. *Papuan languages and the Trans New Guinea family*. Canberra: Pacific Linguistics.

Richert, Ernest and Ttopoqogo Hoopusu. 2002. *Noo supu, a triglot dictionary*. Ukarumpa: Summer Institute of Linguistics.

Ross, Malcolm. 1995. The great Papuan pronoun hunt: recalibrating our sights. In Connie Baak, Mary Bakker, and Dick van der Meij, eds *Tales from a concave world: liber amicorum Bert Voorhoeve,* 139–168. Leiden: Department of Languages and Cultures of South-East Asia and Oceania, Leiden University.

—— 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley et al., eds, 2005, 15–65.

—— In prep. Defining the Trans New Guiena family: preliminary evidence from pronouns. In Pawley et al., eds, in prep.

Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds. 1998. *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society. 1: Material culture*. Canberra: Pacific Linguistics.

Seligmann, Charles Gabriel. 1906. The Hunterian lectures on the physical anthropology and ethnology of British New Guinea (lecture II). *The Lancet* I.425–429.

Smallhorn, Jacinta. In prep. A reconstruction and subgrouping of the Binanderean languages of Papua New Guinea. PhD thesis. Canberra: The Australian National University.

Strong, W.M. 1911. Notes on the languages of the north-eastern and adjoining divisions. *Annual Report for Papua for the Year Ended 30th June, 1911*, 767–781. Printed and Published for the Government of the Commonwealth of Australia by J. Kemp, Government Printer for the State of Victoria.

Suter, Edgar. 1997. A comparative look at the dual and plural forms of verb inflections and pronouns in Northeast New Guinea Papuan languages. *Language and Linguistics in Melanesia*. 28.17–68.

—— 2007. Huon Peninsula etymologies. Unpublished manuscript. Winterthur.

Wilson, Darryl. 1969. The Binandere language family. In *Papers in New Guinea linguistics, No. 9*, 65–86. Canberra: Pacific Linguistics.

—— 1980. English–Suena dictionary, part III. Unpublished manuscript. Ukarumpa: Summer Institute of Linguistics.

Wurm, S.A. 1971. The Papuan linguistic situation. In T.E. Sebeok, ed. *Current trends in linguistics,* vol. 8: *linguistics in Oceania*, 541–657. The Hague: Mouton.

Wurm, S.A., ed. 1975. *New Guinea area languages and language study,* vol. 1: *Papuan languages and the New Guinea linguistic scene*. Canberra: Pacific Linguistics.

Wurm, S.A., C.L. Voorhoeve and Kenneth McElhanon. 1975. The Trans New Guinea Phylum in general. In Wurm, ed., 1975, 299–322.

Z'graggen, J.A. 1980a. *A comparative word list of the Rai Coast languages, Madang Province, Papua New Guinea.* Canberra: Pacific Linguistics.

—— 1980b. *A comparative word list of the Northern Adelbert Range languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.

—— 1980c. *A comparative word list of the Mabuso languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.

—— 1980d. *A comparative word list of the Southern Adelbert Range languages, Madang Province, Papua New Guinea*. Canberra: Pacific Linguistics.

12 *The Papuan languages of the Eastern Bismarcks: migration, origins and connections*

TONYA N. STEBBINS

1 Introduction[1]

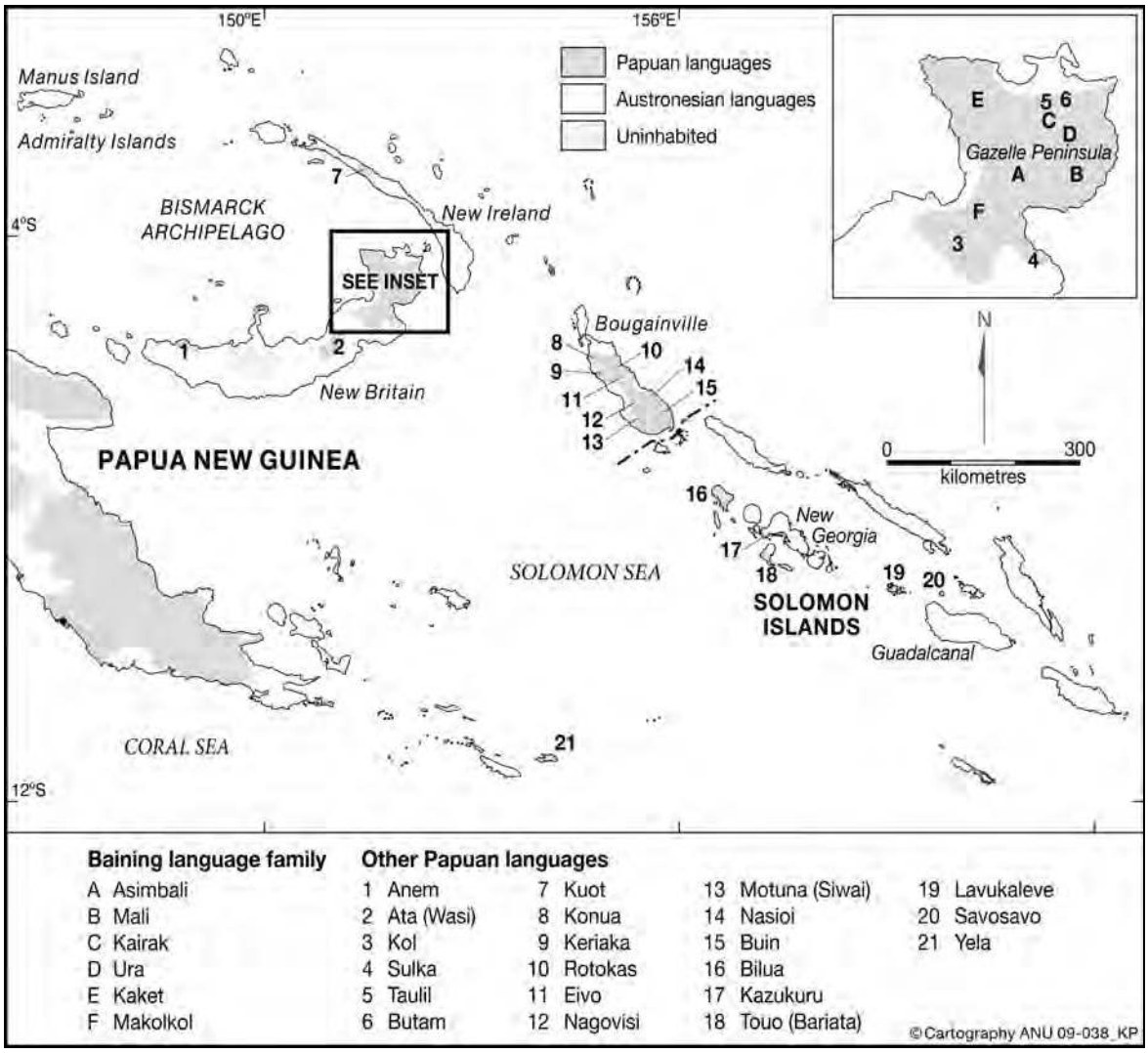
A close reading of early ethnographic sources provides evidence of movements of peoples across the Eastern Bismarcks region that can inform our efforts to understand the linguistic prehistory of this region. This paper compares current proposals for grouping the languages of the region with the information that can be gleaned from early sources and proposes some avenues for future investigation. The paper considers the relationships between the non-Austronesian or Papuan languages of the region in terms of origins and migration, the history of language contact, comparative phonology, gender and noun class marking, pronouns and syntactic structure. This is a preliminary report based on archival evidence as well as new linguistic and ethnographic materials derived from fieldwork in the Taulil, Mali and Simbali (Baining) communities and some more recently published materials by others on Sulka, Kol and Kuot.

The following sections outline the approach (§1.1), summarise two of the more recently proposed groupings of the East Papuan languages (§1.2) and provide further information regarding what is known about the relevant languages (§1.3). The paper then turns to the histories of the communities concerned in §2, focussing on the origins, migration and known patterns of language contact in the relevant communities. Specific characteristics that are associated with the Papuan languages of the Eastern Bismarcks are presented in §3. Finally, a summary of the grouping of the languages suggested as a result of this review, along with some proposals for future research, are set out in §4.

[1] I would like to thank Sheena Van Der Mark, Catherine Easton and Friedel Frowein for discussing this material with me, and also particularly thank Friedel Frowein for a great deal of assistance in navigating the early ethnographic accounts on which much of the basic argumentation of this paper is based. Thanks to Ger Reesink for sharing unpublished data on Sulka and Kol and for his comments on the paper and to Eva Lindström, Bethwyn Evans, Brigitte Pakendorf and the reviewers of the paper for their very helpful comments.

1.1 Scope and aims

The East Papuan group of languages includes the Papuan languages from New Britain right across to the Solomon Islands (see Map 1). Although this term was used in Wurm (1982) to refer to a possible genetic grouping, such a grouping has never been established and the term is now used in a more general way to designate the languages in relation to their geographical position and their non-Austronesian origins.



Map 1: Area map of East Papuan languages

The Papuan languages of the Eastern Bismarcks show evidence of long standing language contact with Austronesian languages. This has resulted in Austronesian influences at all levels of the grammar in these languages and complicates the linguistic history of the region in intriguing ways. Many of the features that distinguish Papuan languages from each other in the Easter Bismarcks appear to indicate different interactions with speakers of Austronesian languages at least as much as they indicate distinct pre-Austronesian genealogical pedigrees.

Although the term East Papuan does not reflect established genetic relationships among these languages, a few areal features are identified in Dunn, Reesink and Terrill (2002) and Terrill in (2002), which focuses on noun classification. In relation to the Eastern Bismarcks specifically, Reesink (2005) provides information about Papuan and Austronesian features in Sulka and its neighbour Kol with some reference to Mali (Stebbins in press) and Kuot (Lindström 2002). In seeking to untangle the linguistic prehistory of the Eastern Bismarcks, the history of language contact must be read in the context of what can be reconstructed about the migrations of various groups through the area, as well as contact between groups of Papuan speakers.

Different types of shared characteristics provide different sorts of evidence of connections. Table 1 summarises the types of characteristics reviewed in this paper and summarises the ways in which it seems most reasonable to interpret them.

Table 1: Evidential status of shared characteristics across languages and speech communities

Characteristic	Evidence of contact	Evidence of genetic relatedness
Shared typological features	Strongly suggestive of contact	Moderately suggestive of relatedness in conjunction with other features
Large numbers of lexical cognates	May indicate more intensive contact	May indicate recent branching
Small numbers of lexical cognates	May indicate more sporadic contact	May indicate distant branching
Shared patterns in paradigms	Parallel but non-cognate paradigms suggestive of language contact	Cognate paradigms strongly suggestive of genetic relatedness
Past or present occupation of contiguous land	Strongly suggestive of language contact	May correlate with genetic relatedness

Table 1 shows that overall evidence of language contact is available from more different sources, and is easier to find, than evidence of genetic relatedness. It also indicates that care needs to be taken in interpreting the presence of cognates, particularly when they are present in small numbers and, consistent with Ross (2001) and Nichols (1996), takes cognate paradigms to be the strongest evidence of genetic relatedness.

The scope of this paper covers all of the Papuan languages currently spoken on the Gazelle Peninsula in East New Britain and in neighbouring areas of the Eastern Bismarcks. These languages include: the Baining languages (Kaket, Kairak, Ura, Mali, Simbali and possibly Makolkol), Taulil, Butam, Sulka, Kol and Kuot (see Map 1). This paper is particularly concerned with the relationship between Taulil, Butam and the Baining languages. In this paper I reconsider this association with reference to early writings about the groups concerned as well as more recently available data.

The main argument of this paper is that a review of the known movements of the various groups concerned does not match either of the groupings that have been suggested for these languages in their finer detail. Such a review is likely to allow us to develop more targeted strategies for exploring relationships among the languages of the area. This paper does not seek to propose a method for substantiating genetic links among all these languages, since such links are not likely to be made, but I am interested in exploring evidence of linguistic influence among different groups insofar as these can be discerned.

1.2 Proposed groupings of the East Papuan languages of the Bismarcks

On the basis of comparison of pronominal forms from across the East Papuan languages, Ross (2001) speculatively proposes the groupings for the non-Austronesian languages as set out in Table 2. (Yele is included in the table on the basis of its apparent relation to the West New Britain languages (Ross 2001:310).) Note that languages listed flush on the left side are considered to be isolates.

Table 2: 'East Papuan' language groupings from the Bismarcks based on pronominal data (adapted from Ross 2001:309).

Yele-West New Britain
Yele
West New Britain
Anêm
Ata (=Pele-Ata = Wasi)
Kol
Sulka
East New Britain
Baining
Taulil
Butam
Kuot

This table does not identify the various Baining languages or make any attempt to distinguish relationships among them and it treats Butam and Taulil as equally distinct from Baining. In many other sources, the distinction between the Baining languages, and between Taulil and Butam is collapsed so that, for example in Ethnologue, we find the arrangement reflected in Table 3. This classification is presumably derived from Wurm (1982).

Table 3: Ethnologue[2] representation of the Yele-Solomons-New Britain group with only the New Britain branch expanded.

East Papuan (36)
Yele-Solomons-New Britain (20)
New Britain (12)
Anem (1) [Anem]
Baining-Taulil (7) [Qaqet, Kairak, Mali, Simbali, Taulil, Ura, Makolkol]
Kuot (1) [Kuot]
Sulka (1) [Sulka]
Wasi (1) [Pele-Ata]
Kol [Kol] (Papua New Guinea)

In this table, the Baining languages are distinguished from one another but the assumedly different status of Taulil is implied rather than clearly represented through its inclusion in a list of Baining-Taulil languages. The inclusion of Makolkol here is speculative but possibly correct (see §1.3). Butam is not mentioned in this table but elsewhere in Ethnologue[3] it is clear that Butam is treated as a sister dialect of Taulil.

1.3 Overview of current knowledge about these languages

The only non-Austronesian language family evident in the Eastern Bismarcks is the comprised of Baining languages. To date, only Mali has been thoroughly documented (Stebbins in press, Stebbins with the assistance of Tayul in press-a, and b). Published data on Baining languages other than Mali remains scarce at this stage. Early descriptions of the Baining languages, with a strong focus on Kaket, are listed in Carrington (1996). Many of the materials recorded for the Baining languages, including manuscript grammars and dictionaries, have not been traced and were probably lost during the Second World War. In the early ethnographic writings of the German Mission, Kaket (called simply Baining) is the language that received the most attention (for example Rascher 1904a). It is also described in a sketch grammar by Parker and Parker (1977). More recently, Stanton (2007) provides an analysis of Ura phonology and morpho-phonology and her examples allow preliminary consideration of a range of other issues. A description of Kairak is currently being developed by Cindy Schneider at the University of New England.

The inclusion of Makolkol with the Baining languages is problematic as there is no published language data on which to base an assessment. However, reports of speakers of Makolkol continue to come in (Ken Sumbuk pers. comm.) so it may eventually be possible to remedy this situation. Early associations of the Makolkols with the Bainings, for example Laufer (1955:37), appear to be based on certain aspects of Makolkol material culture, such as house design (see also Feinberg 1959).

[2] http://www.ethnologue.com/show_family.asp?subid=91731 (accessed 23/07/2008)
[3] http://www.ethnologue.com/show_language.asp?code=tuh (accessed 23/07/08)

In order to properly settle the question of the relationship between Taulil and Butam and the Baining languages, in addition to amassing sufficient reliable data on each of the Baining languages and Taulil and Butam, it will be necessary to reconstruct Proto Baining and then seek to establish a relationship between Taulil and Butam and the ancestor to the Baining languages. One of the difficulties with substantiating the relationship between Taulil and Butam and the Baining languages has been the relative paucity of data available on the languages involved. Although they neighbour the one of the most well developed areas of Papua New Guinea, and one of the most well studied communities, the Tolais, the Baining languages and Taulil and Butam are only now beginning to be more comprehensively described.

It is important to keep in mind that earlier research on the languages of the Gazelle Peninsula has historically occurred in a context in which the Tolai community dominated outsider perceptions of others in the region. This situation and its present day resonances are summarised in Rohatynskyj (2005). This prominence of the Tolais in relationships with outsiders facilitated a strong tendency to lump together the non-Tolai groups of the Gazelle. It seems likely that the use of the term Baining has often been used to refer to the 'non-Tolai' rather than necessarily asserting membership in the Baining ethnic or linguistic group.[4] This may have had a strong influence on the decision to group Taulil and Butam with the Baining languages in previous studies. Substantial linguistic evidence for this connection has yet to be established.

The first data on Taulil were collected by Futscher in the early 1900s. Laufer's (1950) overview of Taulil grammar is based on Futscher's data and, together with Laufer (1959a), is the only fully reliable source of data on the language. An additional wordlist was collected by Neuhaus (n.d.) and in the early 1980s a Summer Institute of Linguistics team prepared a brief write-up of Taulil (Lindrud and Nicholson n.d.). Since 2001, I have made several brief visits to the Taulil community and since 2007 colleagues have also conducted fieldwork there.[5] Together these materials now comprise a wordlist of 1000 entries and a small corpus of texts, augmented by elicited paradigms of various grammatical categories.

Language data on Butam were collected by Futscher in 1936. Extracts from Futscher's notes are published as Laufer (1959b). The available information on Butam includes a fourteen page summary of the grammar that contains roughly 200 different lexical and grammatical words. There is little hope of discovering anything new about Butam since the last fluent speaker of the language is reported to have died in 1938 (Laufer 1959b:193).

An important historical fact that has a bearing on the usefulness of seeking to affirm or deny the association of Taulil and Butam with the Baining languages is the fact that speakers of Taulil and descendants of Butam speakers claim to have migrated in relatively recent times from Southern New Ireland. Their history of migration is discussed in more detail in §2. The migration pathway reported for these two groups suggests that connections between Taulil, Butam and the non-Austronesian languages spoken or

---

[4] See for example the dictionary of the Tolai language published by Bley (1900) whose entry for Butam (cited in Laufer 1959b:186) reads 'Name für die in der Mitte der Gazelle südlich vom Varzin wohnenden Baininger.' [Name for the Bainings living in the middle of the Gazelle in the south of Varzin. - TS]

[5] I would especially like to acknowledge the efforts of Taulil speakers Louisa Gulua and Mary Rickie in collecting material on their own initiative and in partnership with us. Thanks also to all the members of the Taulil community who have contributed stories and other information. Thanks to Sheena Van Der Mark, Mark Lawless, and Cindy Schneider for their assistance in sustaining these relationships and this work while I have been unable to visit the community myself.

originating in New Ireland also need to be considered more carefully. The two relevant languages are Kuot and possibly Sulka. Both of these languages have been the subject of more recent descriptive research. Detailed information on many aspects of Kuot grammar as well as an extensive amount of vocabulary are available in Lindström (2002), while a similar source for Sulka is Reesink (2005) (though the latter is more restricted in scope). Grammar sketches by SIL translators are also available for both languages (Chung and Chung 1996, and Tharp 1996).

Although little is known about Kol (some information is presented in Reesink 2005), it is a close neighbour to the Baining languages and must also be examined for evidence of possible links to this group.

2 Origins, migration and language contact

A good deal of information about the non-Austronesian populations of New Britain and New Ireland is available in historical and ethnographic accounts by Europeans from the turn of the 20th century. A great deal more information is available from speakers of each of the languages concerned. The Bainings are the only ethnic group on the Gazelle Peninsula to claim that they are indigenous to the area (see for example Rascher 1904a:31). All other groups are thought to have migrated from New Ireland. A concise summary of the probable migrations of non-Austronesian speaking groups from New Ireland to New Britain is found in Laufer (1950:629 and fn.5). Laufer states that the Taulil, Butam and Sulka groups are all migrants from Southern New Ireland and that both the Sulka and Butam report that the spirits of their dead return to this area. This general understanding is also shared by present day members of the Taulil community.

In this section I review the historical social relationships between the speakers of the relevant groups, identifying connections that are likely to, or can be shown to have lead to borrowing. The section begins with the history of the Taulil and Butam (§2.1) followed by the Bainings in (§2.2), then the Sulka (§2.3), Kuot (§2.4) and what little is known about the speakers of Kol (§2.5).

As background to this discussion, it is useful to keep in mind that the locally dominant Tolai community, made up of speakers of Kuanua, Vinitiri and Bilur, are thought to have migrated from New Ireland in relatively recent times (Fajans 1997:33, Neumann 1992:237), although precise dates are impossible to establish.

2.1 Taulil and Butam

Sources are consistent in identifying the homeland of the Taulil as being in the vicinity of Muliama, on the East Coast of New Ireland (Laufer 1950:629, William Puongo and Louisa Gulua pers. comm.). According to Puongo (pers. comm.), the Taulil began their migration from Muliama by walking to the west coast, on the other side of the island from Namatanai. Then they travelled by canoe across the Bismarck Sea to Weberhafen (now also known as Ataliklikun Bay), landing first at Epelik, between the Keravat and Tivanakot rivers. In contrast, the Butam are said to have travelled across St George's Channel to the mouth of the Warangoi River and to have moved inland from there, eventually meeting up with the Taulil. Laufer's (1950:629) report appears to have conflated the pathways of the Taulil and Butam groups and reflects the story remembered for the Butam community. All these sources are consistent in asserting that Taulil and Butam have adjacent homelands in

New Ireland in the vicinity of Muliama, particularly to the south and west. The Taulil are apparently no longer remembered in New Ireland but there continue to be people who identify as Butam in the area on the west coast and there are stories about Butam living in the mountains along the centre of the island (Ed Condera and Friedel Frowein pers. comm.). This is the area from which the Butam are said to have embarked for New Britain. The presence of a Butam community in southern New Ireland is also reported by Schlaginhaufen (1908) (see also Friederici 1912:290). He identifies the Butam as living south of the Danfu River (which drains to the East coast, south of Muliama near the Manga Mission) and mentions the villages of Maletambit and Kan. It seems reasonable to speculate that the Butam migration in to New Britain occurred significantly later than the Taulil migration since the latter are no longer remembered in New Ireland.

Both the Taulil and Butam communities were under persistent attack from their neighbours in the period before European colonisation. By 1884, this had resulted in the assimilation of the remaining members of the Butam community into the Taulil community (Laufer 1959:190). The dominance of neighbouring groups continues today. The main influence these days is from Kuanua as speakers of this language have gradually occupied the area and intermarried into the Taulil community. A number of early reports (eg. Kleintitschen 1906:164) noted similarities between Taulil and Kuanua. These similarities appear to be based on intensive contact over several generations. This is reflected, for example, in the phonology of Taulil which is very similar to Kuanua.

2.2 Baining languages

The Baining languages collectively show clear signs of intensive contact with speakers of Austronesian languages at an earlier stage in their history. In more recent times, interactions between speakers of individual Baining languages and their Austronesian speaking neighbours, particularly the Tolais, have varied in intensity but have consistently been shaded by a profound imbalance in social status. This has been most well documented in relation to the Kaket-speaking 'North-Bainings' who were frequently captured by Tolais and killed or taken as slaves; practices that continued into the 20th century (see for example Kleintitschen (1906:11) and Meier (1914:350) who notes that no Tolai person has ever been known to learn the Baining language).

The Central- and Southern-Bainings (the Uras, Kairaks and the Malis and Simbalis), were not the target of early missionisation or of raids, being too far distant from the Tolai community to attract their attention. The Uras and Kairaks currently live in close proximity to the Taulil while the Mali and Simbali communities have historically had more intensive relations (and hostilities) with the Sulka speaking community. Further research examining these languages to measure the extent of borrowings among them will prove interesting.[6]

Given the consistency of the typological patterns shared across the Baining language family (some of which are mentioned below), it seems unlikely that these characteristics are the result of recent Tolai influence on the Baining languages individually since the interactions between each Baining group and the Tolai community have been so different.

---

[6] Reesink (2005:153–155) identifies a handful of shared forms.

2.3 Sulka

The origins of the Sulka speaking community are not clear. A range of hypotheses are examined by Laufer (1955:37–41), who notes that some Sulka describe themselves as coming from the west, and spends considerable space exploring the historical relationship between the Mengen and Sulka communities. Reesink mentions that 'the form and function of Sulka's articles and its attributive adjectival construction appear to corroborate a New Ireland origin' (2005:190). Parkinson (1999:77) notes that the Sulka share a tradition of bride isolation with groups inhabiting the Rossel Mountains in New Ireland.[7] Both these statements correlate with Laufer's report (1950:62:fn.5) that the Sulka themselves claim to have migrated from southern New Ireland (see also Kleintitschen 1906:29). If the Sulka homeland was in this area, this would place them just to the southwest of the homeland of the Taulil and Butam groups.

The long standing and friendly relations between the Mengen and the Sulka communities (Rascher 1904b:209) are reflected in evidence of language contact between Mengen and Sulka reported by Reesink (2005). More recent and relatively intensive interaction between Mali and Sulka has resulted from Sulka migrating into coastal areas along St George's Channel (see Burger 1913:4) at the same time as increasingly permanent Mali settlements were established along the coast in this same area.

As Reesink (2005) demonstrates, Sulka speakers, like their Baining neighbours, have had long term contact with speakers of Austronesian languages. Although the balance of Austronesian and Papuan features is similar to the balance in the Baining languages, the distribution of features is not identical. For example, the realis/irrealis distinction is clearly more prominent in Sulka, where it is incorporated into the forms of the subject proclitics, than in Mali, where it occurs as a clause initial discourse particle. Similarly, although Sulka and Mali share the characteristic of having pre-nominal articles, they differ in the distribution of demonstratives. In Sulka these are pre-nominal but in Mali they follow the NP head.

2.4 Kuot

The Kuot are located further to the north (in New Ireland, just south of the third parallel (Lindström 2002:30)). Lindström (2002:29) reports that at the time of contact with Europeans Kuot villages were located on the western coast of New Ireland and in the mountains. Kuot villages on the East coast have only been established in more recent times. This places the Kuot at some distance from the homeland of the Taulil and Butam and, like the linguistic evidence set out below, suggests that there is no reason to look for or assume close connections between these groups. More distant connections between Kuot, Taulil, Butam and Sulka may eventually be established.

2.5 Kol

In her linguistic bibliography of the New Guinea area, Carrington (1996) refers to scattered wordlists on Kol. Stellan Lindrud from SIL has worked for many years in the community but there has not yet been an in depth description of the language. The most

[7] These mountains apparently form part of the Hans Meyer Range, roughly in line with the Duke of York Islands.

detailed information currently available is in Reesink (2005). Although this paper focuses on Sulka it contains many references to the lexicon and typology of Kol. No information about the history of the community is currently available but Kol is relevant to this discussion as a present day neighbour of Sulka and because given its geographical setting it seems the most likely potential relative to the Baining languages.

3 Characteristics shared by the Papuan languages of the Eastern Bismarcks

This section provides an overview of features that are shared among the Papuan languages of the Eastern Bismarcks. These features are placed into three groups. Features shared with Western Oceanic languages spread widely across the region are presented in §3.1. Characteristics that are more strongly centred on the Eastern Bismarcks are discussed in §3.2. Finally, features generally restricted to Papuan languages in the area are presented in §3.3. Since we are still some way from being able to search among reasonably complete descriptions of the languages concerned in order to establish possible cognates, this approach at least provides us with some indication of where it may be fruitful to focus such a search.

In this section I make use of Lynch, Ross, and Crowley (2002) and Ross (1996:186) in identifying features distributed among Oceanic languages. This discussion also draws on the work of Dunn, Reesink and Terrill (2002), but has a different orientation since the focus is specifically on the languages of the Eastern Bismarcks rather than on the region as a whole.

3.1 Characteristics spread widely across the region

Characteristics spread across the Oceanic languages of the region that are also found to some extent in the Papuan languages of the Eastern Bismarcks include: (A) an inclusive/ exclusive distinction, (B) the use of articles and post-nominal demonstratives, (C) sensitivity to a distinction between alienable and inalienable possession.

Given the strength of these characteristics in the Oceanic languages of the area, it seems reasonable to attribute their presence in the Papuan languages to contact with speakers of Oceanic languages.

A. An inclusive/exclusive distinction

Ross (1996:186) notes that an inclusive/exclusive distinction is associated with Oceanic languages. The inclusive/exclusive distinction is found in Vinitiri, an Oceanic language on the Gazelle Peninsula (Van Der Mark 2007:66). It is also found in Kol (Reesink 2005:169) and Kuot (Lindström 2002:213). This distinction is not found in any other Papuan languages in the Eastern Bismarcks but Dunn, Reesink and Terrill (2002:40) report that it is quite common in East Papuan languages outside this area.

B. Use of articles and demonstratives

Following Lynch (1998:120), Dunn, Reesink and Terrill (2002:36) note that typical ordering of elements within the NP in Oceanic languages is:

(1) ART N ADJ DEM[8]

It is therefore striking to note that articles are present in the Baining languages and Sulka. In each of these languages the article occurs in pre-head position, consistent with the Austronesian pattern as illustrated in (2). In general terms, articles in both Mali and Sulka contrast common nouns with proper nouns (Sulka also makes a distinction between singular *a* and plural *o*).

(2)	Mali	Sulka	
	kama mēliarka	*o mia ruk*	
	kama mēliar-ka	*o mia ruk*	
	ART parrot-M.SG	ART men DEM	
	'the parrot sp.'	'those men'	(Reesink 2005:170)

Articles are also reported to a more limited extent in Taulil and Butam where they were associated with proper nouns. In Butam two of the forms (masculine *to* (*To Kormi*) and feminine *ja* (*Ja Urapin*)) are clearly related to forms in neighbouring Kuanua (Laufer 1959b:200) and appear to be indicative of recent borrowing—they are strongly associated with names that have been borrowed from Kuanua. There is an additional feminine proper noun marker associated with traditional feminine proper names, *e* which Laufer observes is similar to the proper noun article *e* in Sulka and in Birara [Bilur]. The Taulil proper name articles are *to* (masculine) and *e* (feminine). The feminine proper name article is identical to the more traditional Butam form (Laufer 1950:634).

Demonstratives in the Papuan languages of the area generally occur in post-head position, following the ordering typically found in nearby Austronesian languages. There are interesting exceptions, particularly in Kaket which, as Dunn, Reesink and Terrill (2002:36) note, following Parker and Parker (1977:39), shows the order set out in (3) (where MOD can be unpacked as a relative clause, a modifier NP, or similar).

(3) DEM ART N MOD

This is quite different to the order in Mali, as shown in (4). The demonstrative in Kaket appears to be comprised of the non-present tense form of the verb *lu* 'see' plus a concordial pronoun agreeing with the gender of the head noun.

(4) Mali: *kama vlam avang ama andevaung na-ngēt*
ART pig(s) CN.DIST.DEM REL three.M PREP-3N.III
'these three pigs'

Kaket (i): *lu-nget ama velam ama su-nget ama depkuas*
these-PL.NH ART pig ART black-PL.NH ART three
'these three black pigs' (Parker and Parker 1977:36)

Kaket (ii): *a dulka lucha*
a dul-ka lu-ka
ART stone-M.SG DEM-M.SG
'this stone' (Rascher 1904a:58–59)

---

[8] Abbreviations used in this paper follow the Leipzig glossing rules with the following exceptions: I - class I concordial pronoun; II - class II concordial pronoun; III - class III concordial pronoun; CN - count neutral noun class; LONG - long noun class; MOD - modifier; NH - non-human; REL - relator; SPEC - specifier article.

The form *lu* in Kaket is identical to the non-present tense form of the verb *tlu* ‘see’ in Mali and it is possible that this demonstrative is an innovation in Kaket based on an expression such as ‘see them’. As shown in (4), Rascher provides examples of the alternate order also in Kaket.

C. Marking of possession

Lynch, Ross and Crowley (2002:40) note that sensitivity to a semantic distinction between alienable/inalienable possession is common to a variety of structural types of possessive marking in the Oceanic languages. All of the Papuan languages in the sample for which sufficient data is available (Kaket is the only exception here) show evidence of treating a semantic group of either body part terms or kin terms or both in a distinctive way. Reesink (2005:179) questions whether the special treatment of kin terms in Sulka amounts to an alienable/inalienable distinction in the possession marking system in the language.

A second topic of interest in relation to possession is the ordering of the possessor (G) and the possessed noun (N). The usual order associated with Papuan languages is GN (Ross 1996:186; Dunn, Reesink and Terrill 2002:33). This is generally true whether the possessor is expressed as an NP or as a pronoun attached to the possessor noun. Of the languages examined here, this is the case for Mali (Stebbins in press), Kaket (Parker and Parker 1977:20 and example (225) on p.72), Sulka (Reesink 2005:181, Tharp 1996:117–118) and for the Austronesian language Vinitiri (Van Der Mark 2007:308). Kuot is reported to have possessed possessor (NG) order (Lindström 2002:8). The remaining languages show either order in some circumstances. Reesink (2005:181) reports that both orders are present in Kol. The order GN is associated with kin terms while the order NG is associated with body parts.

Taulil and Butam also show both orders associated with different markers of possession. It is not clear from Laufer’s Taulil examples whether either of these orders should be understood as reflecting an alienable/inalienable distinction since the same noun *kalaka* ‘house’ is used in exemplifying both strategies. In Butam it appears that the order GN is associated with inalienably possessed body parts (Laufer 1959b:202–203) and that the use of the possessive marker is not required for all body part terms. Examples are provided in (5).

(5)

Taulil GN	Butam GN	and
loka va kalaka	*ngana ir tamua*	*ngana nok*
man POSS house	1SG.POSS POSS body	1SG.POSS hand
‘the man’s house’	‘my body’	‘my hand’
Taulil NG	Butam NG	
kalaka ito loka	*kapal ata kekavak-bitam*	
house GEN man	things GEN girl-F.DL	
‘the house of the man’	‘the things of the two girls’	
(Laufer 1950:635)		(Laufer 1959b:203)

3.2 Characteristics strongly centred on the Eastern Bismarcks

Characteristics that occur in both the Papuan and Oceanic languages of the Eastern Bismarcks include: (A) AVO word order with split-S marking, (B) the use of prepositions and post-nominal demonstratives, and (C) clause linking without de-ranking.

Given the strength of these characteristics in the Oceanic languages of the area and their rarity in the Papuan languages of the mainland, it seems reasonable to attribute their presence in the Papuan languages to contact with speakers of Oceanic languages.

A. AVO word order with split-S marking

Dunn, Reesink and Terrill (2002:32) note that the 'normal' constituent order in Papuan languages is verb final and that this order is attested in all East Papuan languages outside the Bismarcks apart from a single language, Bilua, in the Solomons. Lynch, Ross and Crowley (2002:49) note that although word order varies across the Oceanic language family, AVO order is the most widespread. The distribution of this order maps reasonably well onto the territory covered by the East Papuan languages. Within the Bismarcks is the use of AVO constituent order is also found in Papuan languages. The single exception is Kuot (Lindström 2002:15) which is VAO, VS. Also widely attested in the languages of the Eastern Bismarcks is the pattern of split-S (unaccusative) marking associated with stative verbs. This pattern is attested in all AVO languages in the sample except Sulka. Examples from Mali are provided in (6).

(6)

A V O			S_A V		V S_O	
ngu	*nen*	*nge*	*ngua*	*met*	*trong*	*ngo*
1SG.II	ask.NPRES	2SG.III	1SG.I	go.PST	be.stiff.PRS	1SG.III
'I asked you'			'I went'		'I am stiff (with cold)'	

B. Use of prepositions

Prepositions are a recognised feature of Austronesian languages in general and are found in Western Oceanic languages (Ross 1996:186). They are also found in all of the Papuan languages referred to in this survey. Dunn, Reesink and Terrill (2002:33) note a strong correlation in the East Papuan languages more broadly between non-final verb constituent order and the use of prepositions. This is in contrast to the Papuan languages of the mainland which typically have OV word order and postpositions. Languages in the sample vary in the number of prepositions that are present in the system. For example there are thirteen simple prepositions in Mali that combine to make eighteen complex forms, giving a total of thirty-one prepositions in the language. A much smaller set is reported for Taulil. While Laufer (1950:635) only recognises one prepositional form there are likely to have been a few more that were categorised in his analysis as adverbs.

C. Clause linking without de-ranking

Dunn, Reesink and Terrill (2002:36) note that there is a strong correlation between non-final verb constituent order and lack of clause chaining, that is, the use of distinct verb forms in non-final clauses. What is particularly interesting about the languages in the Eastern Bismarcks sample is the fact that there is no evidence of any type of 'de-ranking' on linked clauses that is so often associated with subordination. Instead, all the languages

in the survey (including the Austronesian language Vinitiri (Van Der Mark pers. comm.)) appear to have a single set of criteria for forming grammatically acceptable clauses. For example, they generally require concordial pronouns associated with the predicate, and link clauses via apposition and the use of linking particles. From a semantic perspective, these linkers may be involved in coordination (linking independent clauses) or subordination (linking clauses in which the interpretation of one clause is dependent on another), but semantic dependencies between the clauses are not reflected in syntactic dependencies.

3.3 Characteristics generally restricted to Papuan languages

As Dunn, Reesink and Terrill (2002) have argued, typological characteristics specific to a particular set of languages may provide evidence of earlier language contact in a region. Given the proximity of the Eastern Bismarck Papuan languages, this seems to be a particularly apt interpretation of the data presented here. Further research will be required before more detailed analysis of the lexicon allows us to make judgements about the presence of possible cognates shared among these languages. The three shared typological features presented here are: (A) intervocalic lenition, (B) noun class and gender marking, and (C) quinary-vigesimal numeral systems.

A. Intervocalic lenition

A distinctive feature of the non-Austronesian languages of East New Britain and New Ireland is the lenition of intervocalic voiceless stops. This seems to be a remnant of typological profile of the original languages in the area.

(7) [-voice -cont] → [+voice +cont] / V_V

This phonological rule is currently found in the Baining languages and Kuot. It is not reported for any other languages in the area. See examples in (8).

(8)	Mali	Kaket	Kuot
	alechi[9]	*ngu tav a mung*	[mirjevarıvo]
	a=lek-ki	*ngu tap a mung*	*mi-tie parip o*
	SPEC=hole-F.SG	1SG cut.PRS SPEC tree	3PL-there faeces.NSG 3F.POSS
	‘a large hole’	‘I cut down a tree.’	‘those faeces (of) hers’
		(Rascher 1904a:36)[10]	(Lindström 2002:89)

Ross (1994:558–559) identifies a ‘phonological alliance’ linking Kuot with neighbouring Austronesian languages Lamasong, Madak, Barok, Nalik and Kara based in large part on this pattern and surmises (Ross 1994:566) that this feature has spread from Kuot to its neighbours.

---

9 The digraph <ch> represents a voiced velar fricative in Mali.

10 The application of the rule is blocked on verb initial consonants where this alternation now indicates a contrast in tense. For details see Stebbins (in press, Chapter 4).

B. Gender and noun class marking

Terrill (2002) provides an overview of the distribution of noun classification systems in East Papuan languages. In this section, I provide additional data on the Papuan languages of the Eastern Bismarcks. All of these languages, with the exception of Sulka, have two (M/F) or three (M/F/N) term gender marking systems that are typically indexed outside of the NP (for example in verb agreement). These languages also tend to have evidence of much larger, more complex noun class systems. All of these involve relatively large numbers of distinctive categories that trigger agreement on noun dependents and are, in most cases, related to gender marking outside of the NP.

There appear to be at least three different situations with regards to noun categorisation within the languages of interest here. These categories vary in terms of their semantic and formal opacity across the languages. In some languages the system is productive and predictable while in others it appears to be frozen.

The Baining languages typically have nine noun classes, eight of which distinguish three numbers, singular, dual and plural. In Mali the classes are: masculine, feminine, diminutive, reduced, excised, flat, long, extended and count neutral (see Stebbins 2005 for a complete account of the Mali noun class system and its relation to gender marking in the pronouns). The size- and shape-based noun classes have distinct plural forms while the masculine and feminine noun classes leave non-human nouns unmarked in the plural and mark only human plurals (*-ta*).

These noun classes are emblematic of Baining identity. During an early discussion while on my first fieldwork in the Mali community, people volunteered most of these forms and presented them in paradigms in much the same way as speakers of European languages are able to recite noun declensions. The same classes with cognate markers are identified for Kaket in Parker and Parker (1977:9–12). The two systems are compared in Table 4.

Table 4: Noun classes in Mali and Kaket (singular suffixes).

	Mali	Kaket
Masculine	*lik-ka* ‘younger brother’	*ngiam-ka* ‘younger brother’
Feminine	*asen-ki* ‘knife’	*avis-ki* ‘knife’
Diminutive	*amēng-ini* ‘stick’	*meng-ini* ‘stick’
Reduced	*amēng-ēm* ‘tree stump’	*ameng-em* ‘tree stump’
Excised	*amēng-igl* ‘plank’	*meng-igl* ‘plank’
Flat	*chēseng-vēs* ‘feather’	*qasing-es* ‘feather’
Long	*amēng-vēt* ‘pole’	*meng-it* ‘pole’
Extended	*amēng-ia* ‘large log’	*meng-ar* ‘thick post’
Count Neutral	*lat* ‘garden, gardens’	*sleng* ‘garden, gardens’

These noun classes interact predictably with the gender marking system used outside the NP and on possessive pronouns. In Mali, masculine and feminine noun classes correspond to masculine and feminine genders while the other noun classes correspond to the neuter gender.

In Taulil and Butam we also find reports of a range of masculine and feminine noun suffixes as well as suffixes that encode size and shape based distinctions and distinguish singular, dual and plural number. The unmarked plural is even stronger in these languages than in the Baining languages since there is no human plural noun suffix.[11] Masculine and feminine non-human nouns are unmarked in the plural and there are also nouns that do not take noun endings (equivalent to the count neutral class in Table 4). Examples of masculine and feminine nouns in Taulil and Butam are provided in Table 5.

Table 5: Masculine and feminine gender marking in Taulil (Laufer 1950:635) and Butam (Laufer 1959b:635).

Taulil	masculine	feminine	Butam	masculine	feminine
SG	*tul-a* (SG)	*tul-e* (SG)	SG	*pal-a*	*talap*-e
DL	*tup-ip* (DL)	*tul-vitam* (DL)	DL	*pal-ip*	*talap-avitam*
PL	*tul* (PL)	*tul* (PL)	PL	*pal*	*talap*
Gloss	'bird'	'bird'	Gloss	'hut'	'(female) dogs'

The difference between the Baining languages and Taulil and Butam is that in the latter two languages the suffixes seem to be less systematised overall (thus it is not clear whether it is appropriate to assert the languages have a gender system in the fullest sense of the term), and for present day speakers of Taulil they also seem to be less emblematic (at no time has anyone volunteered anything like the paradigms given by speakers of Mali). A similar situation appears to have held for Butam (Laufer 1959b:201–202). Laufer recognises a neuter gender in both Taulil (1950:636) and Butam (1959b:204).

Reesink (pers. comm.) reports that Kol has nine noun declensions. The declension classes include masculine and feminine as well as an additional semantically defined class 'neuter' class referring to pieces of things (apparently analogous to the excised class in Mali). The other classes are reportedly distinguished on formal grounds. Inanimate plural nouns are treated as a distinct class. This is perhaps analogous to the treatment of non-human plural masculine and feminine nouns in Mali. The noun declension appears to be covert (not marked on the noun itself) for all classes except the neuter class which takes the prefix *te-*. Each class is represented in the pronoun system.

(9) Kol (i) feminine class: *ral* 'woman' (pronoun = *no*)
(ii) neuter class: *te-bu* 'piece of pork' (pronoun = *ma*)

Reesink (2005:173–175) describes nine Sulka noun classes (first described by Schneider (1942) see also Tharp 1996:162–163)). The declension classes are based on unmarked singulars with various strategies for plural formation including suffixation, vowel changes and suppletion. An example is provided in (10). Sulka is unique among the Papuan languages of the Eastern Bismarcks as it does not have a gender system. Reesink does not comment on the treatment of dual referents with respect to this system.

(10) Sulka, the suffix *-l*: *a patgie* (SG) *o patgie-l* (PL) 'scraper' (Reesink 2005:173)

---

[11] Instead, there are suppletive forms for human plurals: *lul-a* {man-M.SG}, lul-ip {man-M.DL} lokova 'men' and *lul-e* (woman-F.SG) *lul-bitam* (woman-F.DL) *vakova* 'women'.

For Kuot, the declension classes of nouns identified by Linström (2002:153ff.) are somewhat similar to those identfied by Reesink for Sulka. They are generally formally defined and the semantic unity of the declension classes varies. Singular nouns in most declensions must drop segments before the plural marker is added. Whereas the singular is unmarked in Sulka, the plural seems to be less marked in Kuot since it shows much less formal variation that the singulars do. The plural forms are often variations on *Vp*. Kuot does have gender, a masculine/feminine distinction in the singular, and Lindström (2002:153ff.) provides details about how the noun declensions map onto the gender system. Dual in Kuot is generally but not always built off the plural form by addition of the suffix *-ien*.

(11) Kuot, the *bu* declension: *liobu* (SG) *liobu-p* (NSG) *liobu-p-ien* (DL) 'hole' (Lindström 2002:157)

the *bun* declension: *makabun* (SG) *makauləp* (NSG) *makabie* (DL) 'woman' (Lindström 2002:155)

Given the situation in the Baining languages and in Taulil and Butam, it appears as if there is some sort of correlation (perhaps only a subjective one) between the presence of an unmarked noun class, unmarked non-human plurals and a neuter gender. Lindström (2002:175) refers to Kuot nouns that are non-singular only. She notes (pers. comm.) that when no singular form is known speakers are not able to assign the noun to a declension class.

Whatever the properties of the current systems, the similarities raise questions about possible relations between them. In exploring this area further, scenarios involving both divergence of systems and convergence of systems would need to be considered. The Baining noun class system is regular, morphologically straightforward and marked by substantial phonetic sequences (some noun class markers are two syllables in length) and reasonably consistent across the languages in the family, suggesting that is has a relatively recent origin. Three key questions that should be addressed in a more explicit account are:

1. whether the presence of unmarked plurals is an innovative feature (for example, in the Baining languages, Taulil and Butam),
2. whether the consistent marking of plural and dual in the context of a range of markers for the singular is an innovation (as could be the case for Kuot), and
3. whether or not the relatively opaque declension systems of Sulka and Kuot can be related to the noun class marking system found in the Baining languages.

It is noteworthy that the numbers 'one' to 'three' in Mali have stem modifications to reflect the noun class of the noun. For example the number 'one' has the following forms depending on the noun class of the head noun: masculine and count neutral *asēgēk,* feminine *asēnggik,* diminutive *asēgēni,* reduced *asēgēgl,* flat *asēgēvēs,* excised *asēgēm,* long *asēgēvēt,* and extended *asēngit*. Examples from the functionally unmarked masculine series are provided in (12). Interestingly, Parker and Parker (1977:14–15) report a similar pattern for the numbers 'one' and 'two' in Kaket while Reesink (pers. comm.) notes a numeral classifier system in Kol applying to the numbers 'one' to 'three' that corresponds fairly well with the declension classes found there. These data suggest that noun classes have similar domains of influence across these languages.

C. Quinary vigesimal numeral systems

Lynch, Ross and Crowley (2002:72) reconstruct the numerals from one to ten in Proto Oceanic but note that the decimal system seems to have been falling out of use in some languages that show evidence of earlier (Austronesian) quinary systems. This is true, for example, of Tolai. Quinary vigesimal numeral systems, that is, base five systems with an expression translating as 'whole body' used to represent 'twenty', are also evident in the Papuan languages of the Eastern Bismarcks. Such a system is most likely to have been inherited through the Papuan languages of the area. It is certainly in place in Kol (Reesink pers. comm.), Mali (Stebbins in press), Taulil (Laufer 1950:638), Butam (1959b:209), and Sulka (Reesink 2005:183). It is not fully clear that this system is present in Kaket. Parker and Parker (1977:15) refer to the system making use of 'hand' for the number 'five' and Rascher indicates that 'ten' is expressed by referring to 'two hands' but the form of the number 'twenty' is required in order to be completely certain and is not reported in either source. Lindström (pers comm.) notes that the system in Kuot is decimal. The quinary-vigesimal numeral system in Mali is illustrated in (12).

(12) Mali numerals:

1 *asēgēk* {*a=sēgēk* SPEC=one.M}

2 *aungiom* {*a=ungiom* SPEC=two.M}

3 *adēvaung* {*a=dēvaung* SPEC=three.M}

4 *alevavet* {*a=levavet* SPEC=four.LONG.SG}

5 *angēthikvet* {*angē=tik-vet* 3POSS=hand-LONG.SG}

6 *angēthikvet da sēgēk* {*angē=tik-vet da sēgēk* 3N.POSS=hand-LONG.SG and one.M}

10 *angēthikisem* {*angē=tik-isem* 3N.POSS=hand-LONG.DL}

11 *angēthikisem da sēgēk* {*angē=tik-isem da sēgēk* 3N.POSS=hand-LONG.DL and one.M}

15 *angēthikisem da alēcharavet* {*angē=tik-isem da a=lēchar-vet* 3N.POSS=hand-LONG.DL and SPEC=leg-LONG.SG}

16 *angēthikisem da alēcharavet da sēgēk* {*angē=tik-isem da a=lēchar-vet da sēgēk* 3N.POSS=hand-LONG.DL and SPEC=leg-LONG.SG and one.M}

20 *arucha ma temga* {*a=ru-ka ma tem-ka* SPEC=person-M.SG REL_2 whole-M.SG}

56 *aru atemiom da ngēthichisem da alēcharavet da sēgēk* {*a=ru a tem-iom da angēt= tik-isem da sēgēk* SPEC=person SPEC=whole-M.DL and 3N.POSS=hand-LONG.DL and SPEC=leg-LONG.SG and one.M.SG}

4 Conclusion

On the basis of this discussion, I suggest here a working hypothesis for revised (possibly genealogical) grouping of the non-Austronesian languages of the Eastern Bismarcks as shown in Table 6.

Table 6: Proposal for revised grouping of New Britain/New Ireland languages providing working hypothesis for future research

Central and Eastern languages
Kol
[Baining: Kaket, Kairak, Mali, Simbali, Ura, [Makolkol?]]
New Ireland origin languages
Kuot
[Taulil, Butam]
Sulka

Knowing that Taulil and Butam have come from New Ireland seems like sufficient evidence to allow us to abandon the practice of grouping them with the Baining languages unless strong positive evidence for this relationship comes to light in the future. Present day similarities between Taulil and Butam and the Baining languages could also be contact induced, either as typological features spread across a much wider area or as features spread since speakers of Taulil and Butam began to have contact with the speakers of Baining languages after their migration from New Ireland. It will also be important to keep in mind the fact that the Austronesian influences apparent in Taulil and Butam, for example the proper noun articles, could easily have entered these languages in the period before migration to New Britain since the Austronesian languages of the Gazelle Peninsula are immigrants from New Ireland just like Taulil and Butam.

Based on the historical information and linguistic data outlined above, the most sensible strategy for further research in the area would be to check for possible connections between the Baining languages collectively and Kol, and between Taulil, Butam and Sulka as well as Kuot. At this stage there is no reason to suppose that Taulil and Butam are genetically related to Kuot or Sulka but further data is required before this can be completely ruled out. Certainly any connection between these New Ireland origin languages is likely to be both ancient and tenuous since the data currently available provide no evidence of a genealogical relationship.

I anticipate that the most profitable avenues of future research include a better understanding of the behaviour of unmarked (plural/neuter) nouns, as well as a good comparative lexicon of the languages concerned. Three further morpho-syntactic issues that have not been addressed in this paper but that were identified as potentially significant in Dunn, Reesink and Terrill (2002) are the behaviour of attributive adjectives, the ordering, semantics and dependency of verbal morphology, and the positioning of the negative adverb. These categories are likely to be significant in our developing understanding of the relationships among these languages as the relevant data becomes available.

References

Bley, B. 1900. *Woerterbuch der Neu-Pommerschen Sprache (kanakisch-Deutsch).* Muenster: Tumbrink.

Burger, F. 1913. *Die Küsten- und Bergvölker der Gazellehalbinsel. Ein Beitrag zur Völkerkunde von Neuguinea unter besonderer Hervorhebung rechtlicher und sozialer Einrichtungen.* Stuttgart: Verlag von Strecker und Schröder.

Carrington, Lois. 1996. *A linguistic bibliography of the New Guinea area.* Canberra: Pacific Linguistics.

Clifton, J. ed. 1996. *Two non-Austronesian grammars from the Islands.* Ukarumpa: Summer Institute of Linguistics.

Chung, C-H and K-J Chung. 1996. Kuot grammar essentials. In Clifton, ed., 1996, 1–75.

Dunn, Michael, Ger Reesink and Angela Terrill 2002. The East Papuan languages: a preliminary typological appraisal. In *Oceanic Linguistics* 41(1).28–62.

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

Fajans, J. 1997. *They make themselves: work and play among the Baining of Papua New Guinea*. Chicago: The University of Chicago Press.

Fienberg, David M. 1959. Meeting the Mokolkols. *The Bulletin* 6/5/59, 19,45; 13/5/59,19,45; 20/5/59, 19,45; 27/5/59, 19,82; 3/6/59,19.

Friderici, G. 1912. Wissenschaftlische Ergebnisse einer amtlichen Forschungsreise nach dem Bismarck-Achipel im Jahre 1908, II: Beitraege zur Voelker- und Sprachenkunde von Deutsch Neuguinea. *Miiteilungen aus dem Deutschen Schützgebieten, Ergänzungsheft 5.*

Kleintitschen, A. 1906. *Die Küstenbewohner der Gazelle Halbinsel, ihre Sitte und Gebräuche.* Hiltrup-bei-Münster: Herz-Jesu-Missionshaus.

Laufer, Carl. 1950. Die Taulil und ihre Sprache auf Neubritannien. *Anthropos* 45.627–40.

—— 1955. Aus Geschichte und Religion der Sulka. *Anthropos* 50.32–64.

—— 1959a. P. Otto Futscher, M.S.C., Taulil-Grammatik und naturwissenschaftliche Sammelarbeiten (Neubritannien, Südsee). *Anthropos* 54.213–217.

—— 1959b. P. Futschers Aufzeichnungen über die Butam Sprache (Neubritannien). *Anthropos* 54.183–212.

Lindrud, S. and R. Nicholson. n.d. Notes on the Taulil language of East New Britain. Unpublished manuscript. Ukarumpa: Summer Institute of Linguistics.

Lindström, Eva. 2002. Topics in the grammar of Kuot: a non-Austronesian language of New Ireland, Papua New Guinea. PhD thesis. Stockholm University.

Lynch, John. 1998. *Pacific languages: an introduction.* Honolulu: University of Hawai'i Press.

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

Meier, J. 1914 Review of Burger (Die Kuestenbewohner...). *Anthropos* 9, 348–352.

Neuhaus, K. n.d. *Woeterbuch und Beispielsammelung der Butam-Sprache.* Typescript 100p.

Neumann, K. 1992. *Not the way it really was: reconstructing the Tolai past.* Honolulu: University of Hawaii Press.

Nichols, Johanna. 1996. The comparative method as heuristic. In Mark Durie and Malcolm Ross, eds, 1996, 39–71.

Parkinson, Richard. 1999. (1907.) *Thirty years in the South Seas.* Honolulu: University of Hawai'i Press. Originally published as *Dreissig Jahre in der Südsee.* Stuttgart: Strecker and Schröder. Translated by J. Peter White.

Parker, J. and D. Parker 1977. Baining grammar essentials. Unpublished manuscript. Ukarumpa: Summer Institute of Linguistics.

Rascher, M. 1904a. Grundregeln der Bainingsprache. *Mitteilung des Seminars fürOrientalishce Sprachen zu Berlin.* 7(1).31–85.

—— 1904b. Die Sulka. *Archiv für Anthropologie* 1(4).209–235.

Reesink, Ger P. 2005. Sulka of East New Britain: a mixture of Oceanic and Papuan traits. *Oceanic Linguistics* 44(1).145–193.

Rohatynskyj, M.A. 2005. On knowing the Baining and other minor ethnic groups of East New Britain. In Pamela J. Stewart and Andrew Strathern, eds *Anthropology and consultancy: issues and debates,* 25–45. New York: Berghahn Books.

Ross, Malcolm. 1994. Areal phonological features in north central New Ireland. In Tom Dutton and Darrell T. Tryon, eds *Language contact and change in the Austronesian world,* 551–572. Berlin: Mouton de Gruyter.

—— 1996. Contact-induced language change and the comparative method: cases from Papua New Guniea. In Mark Durie and Malcolm Ross, eds, 1996, 180–217.

—— 2001. Is there an East Papuan phylum? Evidence from pronouns. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

Schlagingaufen, O. 1908. Reisebericht aus Süd-Neu-Macklenburg. *Zeitschrift für Ethnologie* 40.566–567.

Schneider, J. 1942 Grammatik der Sulka Sprache. Unpublished manuscript. Kokopo: Archdiocese, M.S.C.

Stanton, Lee, 2007. Topics in Ura phonology and morphophonology, with lexicographic application. MA thesis. Christchurch: University of Canterbury.

Stebbins, Tonya N. 2005. Nominal classification in Mali. *Anthropological Linguistics* 47(1).77–131.

—— in press. *Mali (Baining) grammar*. Canberra: Pacific Linguistics.

Stebbins, Tonya N. with the assistance of J. Tayul. in press-a. *Mali (Baining) dictionary.* Canberra: Pacific Linguistics.

—— in press-b. *Mali (Baining) texts.* Canberra: Pacific Linguistics.

Terrill, Angela. 2002. Systems of nominal classification in East Papuan languages. *Oceanic Linguistics* 41(1).63–88.

Tharp, Doug. 1996. Sulka grammar essentials. In Clifton, ed., 1996, 77–179.

Van Der Mark, Sheena. 2007. A grammar of Vinitiri: an Austronesian language of Papua New Guinea. PhD thesis. La Trobe University.

Wurm, S.A. 1982. *Papuan languages of Oceania*. Tubingen: Narr.

Part II
Historical development of languages across time

13 *On the zero (voice) prefix and bare verbs in Austronesian languages of Nusa Tenggara, Indonesia*

I WAYAN ARKA

1 Introduction[1]

A zero prefix has no phonological material but is considered present on the basis of functional and paradigmatic opposition in a particular grammatical system. Consider the following different forms of the verb 'collide' in Balinese.[2]

(1) Voice paradigm with the verb *palu* 'collide' (Balinese)

a.	*m-(p)alu*	(<N-*palu*)	'AV-collide'
b.	*Ø-palu*		'UV-collide'
c.	*ka-palu*		'PASS-collide'
d.	*ma-palu*		'MV-collide'

The bare form *palu* (1b) is used to express the Undergoer Voice (UV) in Balinese. It can be analysed as having a zero prefix, represented by *Ø-*, on the basis of systematic formal opposition with the other forms in (1). The zero prefix is a marker prefix in the same morphological slot as other voice markers in Balinese grammar, namely, the Actor Voice (AV) prefix *N-*,[3] the passive (PASS) prefix *ka-*, and the middle voice (MV) prefix *ma-*. In this view, all verb forms in Balinese are equally marked to encode their specific voice types (Arka 2003).

[1] The research reported in this paper was supported by the NSF grant BCS-0617198, where the author is a CI. The data on Rongga was also collected as part of the Rongga Documentation Project funded by the ELDP grant (IPF0011) awarded to the author. I thank the following people who helped with data collection: Ivan Ture (Rongga), Hanafi (Sasak), Petrus Pita (Nage), Fransiskus Remba (Ende), Ismail Saleh and Salma (Bima). Special thanks must go to Fay Wouk who let me share her data on Sasak, Sumbawa, and Bima, and who also provided me with useful detailed comments on an earlier version of the paper.

[2] Abbreviations used in this paper: A – actor; APPL - applicative; AV - Actor Voice; DEF - definite; IR - irrealis; EMP - emphatic; LIG - ligature; LOC - locative; MV - Middle Voice; NEG - negator; p - plural; OV - Object Voice; PART - particle; PASS - passive; PERF - perfective; R - realis; REL - relativiser; s - singular; U - Undergoer; UV - Undergoer Voice; 1,2,3 – first, second, third person.

[3] The homorganic nasal prefix *N-* is phonologically conditioned. It replaces the stem-initial consonant and is realised as /m/, /n/, or /ŋ/ depending on the replaced consonant, e.g., *N-+palu* → *malu* as in (1a).

Evidence for the functionality of the zero prefix in Balinese comes from the examples in (2), where the verb with the zero (UV) prefix, *uber* 'UV.chase', is not intersubstitutable with the verb with the nasal prefix, *nguber* 'AV.chase'.[4]

(2) a. *Tiang nguber/*uber Nyoman* (Balinese)
1 AV.chase/UV.verb Nyoman
'I chased Nyoman.'

b. *Nyoman uber /*nguber tiang*
Nyoman UV.chase /AV.chase 1
'I chased Nyoman.'

Both sentences in (2) have the same meaning: 'I chased Nyoman' (i.e., with the first person 'I' being the actor). In (2a) where the Actor is the Pivot/subject,[5] only the AV verb *nguber* can be used. In (2b) where the Undergoer is the subject, only the UV verb can be used. In short, the nasal prefix and the zero prefix are functionally in contrastive distribution in Balinese morphosyntax.

Other Austronesian languages of Nusa Tenggara, east of Bali,[6] also feature bare verb forms. While the analysis of the zero prefix for bare verbs in Balinese is well motivated, extending the same analysis to bare verbs in these languages is problematic because the functional contrast of the type seen in Balinese is lacking.

Consider the data in (3) from the Puyung variety of Sasak. The nasal verbs (*meli, mancing*) and the bare verbs (*beli*, *pancing*) are intersubstitutable. This indicates that, unlike Balinese (cf. example (2)), Sasak shows no functional opposition of the *N-* and *Ø-* prefixes in this context (see further discussion in §2.2).[7]

(3) a. *Amir* ***beli/meli*** *buku jari Tono* (Puyung Sasak)
Amir buy/N.buy book for Tono
'Amir bought a book for Tono.'

b. ***Kanak=nó*** *jangke=**n*** *pancing* / ***m**ancing* *mpaq*
child=that PRES=3 catch N.catch fish
'The child is catching fish.' (Austin 2001, ex. 33)

In what follows, I present further data from the languages of Nusa Tenggara, Indonesia. I examine the functionality of the zero prefix in the context of how coding resources are used to express different realisations of core arguments. Section 2 begins with the definition of core arguments/relations and is followed by a discussion of how coding resources are employed to express these core arguments in Balinese (§2.1), Sasak (§2.2), Sumbawa (§2.3), Bima (§2.4), Manggarai, and Rongga, Kéo, and Ende (§2.5–§2.6). I show that the case illustrated by Puyung Sasak is an indication of the attrition of the distinctive

---

4 The bare verb is also often called the oral verb in contrast to the nasal verb (i.e., that with the nasal prefix).

5 I use the notion of Pivot as described in Foley and Van Valin (1984) and Van Valin and LaPolla (1999), which is basically the same as the surface grammatical subject (SUBJ) as described in contemporary syntactic theories such as LFG (Bresnan 2001).

6 The region of Nusa Tenggara (lit. the southeast islands) in Indonesia includes the islands of Bali, Lombok, Sumbawa, Flores, (West) Timor, Sumba, and other surrounding small islands.

7 Informal Indonesian also shows that the nasal and zero-verbs are not in contrastive distribution as seen in the following example:

Mereka (me-)langgar aturan (informal Indonesian)
they (AV-)violate regulation
'They violated regulations.'

function of Austronesian voice morphology, and that the languages of central Flores such as Keo and Ende exemplify the ultimate loss of the Austronesian voice system. Section 3 provides a summary and discussion of the merit of positing a zero prefix for language-specific analyses and the theoretical issues of such an analysis.

2 Coding resources for core relations

The term 'core relations' is used to refer to the main argument relations associated with a head predicate (typically a verb) that expresses the relational involvement of participants in the state of affairs (states or events) depicted by the predicate. In the terminology commonly adopted in linguistic typology (Dixon 1979, 1994), these are represented as A, O (or P, as in Comrie (1978)), and S, or as Actor and Undergoer in Role and Reference Grammar (RRG) (Foley and Van Valin 1984; Van Valin and LaPolla 1999).

Core relations are associated with one of the basic functions of language in communication, namely, the expression of the distinction between the 'doer' (i.e., the Actor) of an action and the 'undergoer' (i.e., the participant that is acted upon) in a typical highly agentive two-participant event. This basic distinction may be coded in more than one way by verbal marking, dependent marking, and linear order. The Austronesian languages of Nusa Tenggara, Indonesia, show variation in the availability and employment of these strategies. This is discussed below, beginning with Balinese, followed by Sasak, Sumbawa, Bima and the Flores languages.

2.1 Balinese

Balinese employs all of the three strategies to express different core relations: voice morphology, NP versus PP coding for core versus oblique distinction of arguments, and strict constituent order with possible backgrounding of the A in passive. For example, the verb *uber* 'chase' takes two arguments: 'the chaser' (Actor) and 'the thing chased' (Undergoer). Two possible alternative realisations of the argument relations are exemplified in (2) above. In the AV construction (2a), the verb is marked by the prefix *N-*; the A expressed in an NP precedes the verb, and the U, also expressed as an NP, follows the verb. The reverse holds in the UV counterpart (2b): the U NP precedes the verb, and the A NP follows it. In the passive, exemplified in (4) below, the A is backgrounded and demoted to Oblique (expressed by a PP), and the verb should bear passive morphology (*ka-*).[8]

(4) *Nyoman ka-uber teken ipun* (Balinese)
Nyoman PASS-chase by 3
'Nyoman was chased by him/her.'

Further discussion about the precise nature of verbal derivation in Balinese is given in §3.1 and §3.2.

[8] There are two passives in Balinese — the *ka-* passive and the *-a* passive — dubbed the 'high' and 'low' passive, respectively, in Arka (2003) because of the difference in register between them. Syntactically, both show a typical passive structure, e.g., an alternation of SUBJ → OBJ (of the agent) and OBJ SUBJ (of the patient). The bound form *-a* is also a third person clitic appearing to encode an actor of the UV structure; hence, it serves a double life in contemporary Balinese (see Arka 2008 for details).

2.2 Sasak

While Sasak shows significant variation across its varieties,[9] it also exhibits voice morphology, with verbs displaying a distinction between *N-* (nasal) and bare forms. The Ngeno-Ngene dialect of Sasak retains relatively more verbs with possible *N-* forms than other dialects on the island. Even in this dialect, the opposition is not functional in the context where the actor is given prominence, for example, the 'fronted' sentence as initially seen in (5). Either the bare or the nasal verb can be used in this context.

(5) *Aku jengke-ng=ku bace / **m**bace buku=ni* (Narmada Ngeno-Ngené Sasak)
I PROG-LIN=1 Ø.read / N.read book=this
'I am reading this book.'

However, it should be noted that the *N-* versus bare verb distinction in Sasak is not fully neutralised. There is at least one context in which they are not intersubstitutable (discussed further in Arka (2009)). This case occurs when the Undergoer question word *epe*[10] is fronted, as in (6). Only the bare form can be used in this context, as seen by the contrast between (6a) and (6b). Note that this constraint, and hence the contrast shown in (7), is exactly parallel to that found in Balinese where the UV form must be used when the Undergoer is questioned.

(6) a. *epe te Amir **paleng** rubin?* (Selong Ngeno-Ngené Sasak)
what PART Amir Ø.steal yesterday
'What did Amir steal yesterday?'

b. **epe (te) Amir **maleng** __ rubin?*
what PART Amir N.steal yesterday

(7) a. *Apa ane paling cai?* (Balinese)
what REL UV.steal 2
'What did you steal?'

b. **Apa ane cai maling __?*
what REL 2 AV.steal

An important feature that distinguishes Sasak from Balinese is what I call 'double argument expressions'. This is the structure where the Actor of a transitive verb can still appear in the subject position (realised as a clitic) and is then cross-referenced by a post-verbal PP.[11] This is exemplified in (8a). This structure, dubbed the *isiq* construction by Kroon (1998), is syntactically not passive because the Actor argument is still the Subject,

---

9 While a number of dialects have been identified for Sasak (Jacq 1998), there is no clear linguistic evidence for dialect grouping (Wouk pers. comm.). In the absence of good evidence to support dialect identification, the term 'variety,' rather than 'dialect,' is therefore used in this paper, e.g., Ngeno-Ngené variety. In addition, a particular variety is often identified by the locality where it is spoken, e.g., Selong (Ngeno-Ngené) Sasak.

10 Phonetically, this is [əpə] in Selong Sasak and [apa] or [apə] in other varieties (Wouk pers. comm.).

11 The doubling of the bound pronominal on the verb with free-floating PPs/NPs is quite common in the Austronesian languages of Nusa Tenggara, e.g., Sumbawa, Bima (discussed in this paper), and Kamberra (Klamer 1998), as well as other Austronesian languages of eastern Indonesia beyond Nusa Tenggara: it is common in the Austronesian languages of Maluku such as Taba (Bowden 2001), Alune (Florey 2001), and Kei.

appearing in the preverbal subject position as the clitic *=n*.[12] The *isiq* PP is not obligatory, as an adjunct-like unit adding further specification to the referentiality of the Actor. However, Sasak does have a truly passive structure, exemplified in (8b), in which case the verb bears passive morphology (*te-*). Crucially, unlike in (8a), the subject in (8b) is the Undergoer because the *isiq* PP does not cross-refer to the subject. In (8b), the clitic in the subject position is *=k*, cross-referenced by the free sentence-initial pronoun *aku*.

(8) a. Y*aq=n gitaq kanak-kanak=nó isiq Herman* (Puyung Sasak)
fut=3 see reduplicated-child=that by Herman
'Herman will see the children.' (Austin, MLI paper, 2002)

b. *(Aku) wah=k te-empuk isiq Ali*
1s PERF=1 PASS-hit by Ali
'I have been hit by Ali.' (Shibatani 2008, ex. 8)

Shibatani (2008) uses terminology from Philippine linguistics, analysing the bare verb construction in (8a) as Patient Focus (PF) and the construction with the nasal verb as the Actor Focus (AF). He argues that sentence (8a) is a PF structure because the A is backgrounded and the Patient is given prominence (even though in this particular example the Patient also occurs postverbally). Note that what Shibatani calls a PF structure is primarily characterised by the presence of the backgrounded *isiq* PP accompanying the bare transitive verb. However, in the absence of the *isiq PP*, the bare verb is intersubstitutable with the nasal verb, as seen in (5).

AF and PF structures are functionally (and, arguably, grammatically) distinct structures in Sasak that determine the basis for possible relativisation (see Shibatani (2008) for details).

Given the fact that the bare verb itself can appear in AF and PF constructions in Sasak, there seems to be no good reason to posit a zero prefix for the bare verb in Sasak.

It is worth highlighting the point that the PF structure in Sasak is not exactly equivalent to the UV verb in Balinese (which Shibatani also claims is a PF structure). While the Patient in the PF structure in Sasak and Balinese is given prominence and classified as Topic by Shibatani in both languages, the PF structures in these two languages are crucially distinct. The Balinese PF structure is always grammatically classified as UV, where the UV coding is morphologically distinct as discussed above. By contrast, the Sasak PF construction is not, strictly speaking, classified as UV because the actor argument in the Sasak PF structure remains the subject syntactically; that is, it appears in the subject position (cf. the clitic *=n* in (8a)). Given our current understanding of Sasak, it is not clear whether we do want to claim that Sasak has UV at all. The coding of PF is purely constructional. That is, at the level of verbal morphology, the bare verb cannot be exclusively associated with PF.

---

[12] The distribution and forms of subject clitics in Sasak may vary across varieties of Sasak (cf. Austin 2002). The generalisation is that it is a second position clitic, and, in the presence of an auxiliary or a negator before the verb, the auxiliary or the negator would host the clitic. Otherwise, the verb may host the clitic, where it may be encliticised to the verb as in the following example:

Nu ie kanak saq gitaq=kò nuong (Ganti Menu-Meni Sasak)
that 3 child REL see=1s that (Austin 2002, ex. 43)
(a) 'That is the child who I saw.'
(b) 'That is the child who saw me.'

2.3 Sumbawa

On the basis of the available data to date, Sumbawa (Austin 2001, 2002; Shibatani 2008; Shiohara 2000; Wouk 2002) appears to have almost lost the Austronesian nasal voice morphology.[13] As for Sasak, there is no good reason to posit a zero voice prefix in this language, even though it displays a passive-like alternation. As shown by the examples in (9), there is no nasal prefix on the verb of the 'active' (transitive) structure in (9a). Both core arguments (A and P) in Sumbawa, when pronominals (except for the third person, which is zero), are cliticised on the verb. A free NP may co-index a clitic; for example, *aku* co-indexes *=ku* in (9b).[14]

(9) a. *Ka=ku=pukil=mu* (Sumbawa)
PERF=1s=hit=2s
'I've hit you.' (Shibatani 2008, ex. 10b)

b. *Aku na=i=pukul=**ku** ning Ali* (Sumbawa Taliwang)
1s FUT=PASS=hit=1s by Ali
'I will be hit by Ali.' (Shibatani 2008, ex. 10f)

c. *Surat=nan mu=tulis*
letter=that 2s=write
'You wrote the letter.'

d. *Ali* [*adé ka=**ya**=pukel bedus*] *sakit* (Sumbawa Jereweh)
Ali REL PERF=3=hit goat sick
'Ali, who has hit a goat, is sick.' (Shibatani 2008, ex. 49)

The verb can appear in bare form, without bound clitics. In this case, its free core NP arguments appear before and after the verb as exemplified in (10a). The preverbal NP is the understood A, which is also the Topic and Subject. The postverbal NP is the Object.

(10) a. *tau=lokaq beri tòdé=nan*
person=old love child=that
'The parents love the child.' (Shibatani 2008, ex. 22)

b. **tòdé* [*adé tau=lokaq beri Ø*]
child REL person=old love
'the child whom the parents love'

c. *tòdé* [*adé ya=beri ling=tau=lokaq*]
child REL 3=like by=person=old
'the child who is loved by his/her parents' (Shiohara, 2000:88)

Sumbawa appears to have a passive prefix *ka-* as seen in (11a). However, a bare verb can also turn up with a backgrounded PP Agent as in (11b), which is structurally similar to the analytic passive encountered in Bima, Manggarai, and Rongga.

---

[13] The *N-* prefix does occur, but is restricted to intransitive clauses (Wouk 2002). However, it remains to be checked whether this is true across all varieties of Sumbawa.

[14] The gloss PASS for *i=* in example (9b) follows the gloss given by Shibatani (2008). The bound form *i-* in the Jereweh/Taliwang variety might be comparable to *ya* in Sumbawa Besar. The construction of *i=* followed by a backgrounded PP agent in the Jereweh variety might have been grammaticalised to become a passive marker. This looks like the grammaticalisation of the clitic *=a* '3s' to a PASS marker in Balinese (Arka 2003, 2008). However, further evidence is needed to support a definitive analysis for the status of *i=* in Sumbawa.

(11) a. *ka-ajak-ku ling dengan-ku lalo ko Moyo*
PASS-invite-1s by friend-1s go to Moyo
'My friend invited me to go to Moyo.'

b. *Andi pukil ling Iwan*
Andi hit by Iwan
'Andy was hit by Iwan.' (Wouk 2002, ex. 40, 44)

Sumbawa and Balinese show a similar relativisation constraint in that, when the A is expressed as an NP, presumably in the leftmost prominent position of the clause, the P cannot be relativised. The A-V-P structure in (10a) appears equivalent to the Balinese (AV) A-V-P structure given in (12a). Relativising P in this type of structure is prohibited in both Sumbawa and Balinese, as seen by the contrast between (10a) and (10b) (Shibatani 2008) and between (12a) and (12b).

(12) a. *Meme-bapa-n-ne nyayang-in pianak-ne* (Balinese)
mother-father-LIG-DEF AV.love-APPL child-DEF
'The parents love the child.'

b. **Pianak* [*ane meme-bapa-n-ne nyayang-in* __]
child REL mother-father-LIG-DEF AV.love-APPL
(FOR: '(the) child whom the parents love')

c. *Pianak* [*ane ka-sayangin baan meme-bapan-ne*]
child REL PASS-love-APPL by mother-father-LIG-DEF
'(the) child who is loved by the parents'

The similarity between the acceptable sentences in Sumbawa (10c) and Balinese (12c) should also be noted. In both examples, the A arguments are backgrounded and expressed by postverbal free PPs. There is compelling evidence that the Balinese structure is syntactically passive with the A argument being an Oblique (see Arka 2003 for details) and the P the grammatical Subject (and the understood Topic within the relative clause).

However, the structure with backgrounded A in Sumbawa (e.g., as in (10c)) is quite different. Here, the A argument is arguably the Subject because of its coding as a proclitic in the subject position in the clause. This backgrounding of A with a pronominal clitic in the subject position in Sumbawa is therefore like the Sasak construction seen in (8).

A notable difference is that that the Actor in Balinese (12c) is syntactically demoted to oblique status. This structure is syntactically intransitive and passive, as seen from the verbal coding with passive morphology (*ka-*). On the other hand, the backgrounded PP structure in Sumbawa of the type shown in (10c) is syntactically transitive with both A and P being core arguments. The verb has no verbal passive morphology. Hence, we have a case of backgrounding without demotion of the A argument. The same is true for the equivalent *isiq* structure in Sasak.

To sum up, Sumbawa displays no nasal versus zero-prefixed/bare verb distinction. The bare verb may be used in a construction that expresses a highly prominent A NP argument equivalent to the Balinese AV, in which case the relativisation of its P is barred. However, Sumbawa is like Sasak in that the bare verb can also be used in a structure with a backgrounded PP Actor, in which case the Patient is prominent with the Actor remaining in its subject position and realised by a proclitic.

2.4 Bima

Bima (Arka 2000, 2008b; Austin 2001; Jauhary 2000; Wouk 2002) is also an SVO language with the verb possibly taking optional verbal subject agreement. Pronominal agreement on the verb, if present, carries different aspectual meanings/tenses. The prefix is associated with irrealis forms and the suffix with realis forms. Certain bound forms are homonymous, for example, *-ku* 'emphatic/evidential suffix for future certainty' versus *-ku* '1s.Realis.' The suffix requires the presence of the prefix on the verb as in (13a).

(13) a. ***Nahu*** ***ku****-tu'ba-****ku*** *nggomi* (Bima)
1s 1s.IR-stab-EMP 2
'I will (certainly) stab you.' (i.e., a threat)

b. ***Nahu*** *tu'ba-****ku*** *nggomi*
1s stab-1s.R 2
'I stabbed you.'

As in Sasak and Sumbawa, in Bima, there is no good reason to posit a functional zero affix that marks a voice type for the same reason outlined above; namely, the same verb form can be used to encode two grammatically distinct structures. Consider the following examples showing argument alternations in Bima.

(14) a. *Sia doho di kadera* (Bima)
3s sit LOC chair
'(S)he sat on the chair.' (Jauhary 2000, ex. 25a)

b. *Sia doho-kai kadera.*
3s sit-APPL chair
'(S)he sat on the chair.' (Jauhary 2000, ex. 25b)

c. *Kadera ede doho-kai ba sia*
chair that sit-APPL by 3s
'The chair was sat on by him'. (Jauhary 2000, ex. 28a)

The sentence in (14a) is intransitive, with the locative *kadera* appearing as an Oblique marked by the preposition *di*. In (14b), the sentence is the applicative structure with the locative *kadera* promoted to object, expressed as an NP and appearing postverbally. (14b) is an active transitive structure, and (14c) is its passive counterpart. There is evidence that the *ba* construction with a bare verb (i.e., without an Actor clitic) is syntactically passive (Jauhary 2000). Thus, in this analysis, (14c) is intransitive with the backgrounded Actor being an Oblique, prepositionally marked by *ba* 'by'. Crucially, both (14b) and (14c) employ the same verb form, *doho-kai*.

Even when pronominal agreement is present, the same verb form (together with the agreement affix) is used for a fronted Undergoer question. For example, the same verb form, *mpanga-na,* is used both for the structure in which U appears in its canonical postverbal position (15a) and for the structure in which U is fronted (15b). These two structures appear to be the 'active' structures, i.e., the Actor marked by *-na* is the Subject in the same way as *-ku* in *tubaku* in (13b). In other Austronesian languages with a functional AV-UV distinction such as Balinese and Indonesian, the pair of structures equivalent to that in (15) would require obligatory distinct verbal morphology. In the case of a fronted P content question, this should also be accompanied by the removal of the Actor from its subject status.

(15) a. *Mpanga-**na** au awin* (Bima)
steal-3s what yesterday
'What did s/he steal yesterday?'

b. *Au la Ami mpanga-**na** __ awin?*
what ART Ami steal-3s yesterday
'What did Amir steal yesterday?'

Bima, however, has passive morphology: *di-* for 'irrealis passive' and *ra-* for 'realis passive'. In addition, the passive morphology is accompanied by backgrounding of the Actor and demotion of the Actor to Oblique status. Typically, the Actor appears as a postverbal PP, a clear sign of Oblique status, as seen in the following alternation.

(16) a. *Iwa nahu sepe-na buku ede* (Bima)
friend 1s borrow-3s book that
'My friend borrowed that book.'

b. *Buku ede ra-sepe ba iwa nahu*
book that PASS.R-borrow by friend 1s
'The book was borrowed by my friend.'

The passive morphology can also appear with pronominal suffix agreement as in (17). In this case, the Actor can appear as an NP cross-referencing the pronominal suffix.[15] However, unlike an 'active' Actor NP, the Actor can typically no longer appear in a preverbal position, as seen from the following contrast.

(17) a. *Au **ra**-mpanga(**-na**) la Ami awin*?
what PASS.R-steal(-3s) ART Ami yesterday
'What was stolen by Amir yesterday?'

b. **Au la Ami ra-mpanga-na awin?*
what ART Ami PASS.R-steal-3s yesterday

To sum up, there are at least three kinds of passive constructions in Bima: (i) passive without passive morphology where the A argument is expressed postverbally as a PP (not an NP) as in (14c); (ii) passive with passive morphology without pronominal cross-referencing of the Actor on the verb by a PP, as in (16b); and (iii) passive with passive morphology with possible pronominal cross-referencing of the Actor on the verb by an NP (17a). While Bima shows active-passive alternations, the same bare verb can be used in both the active and passive structures; hence, no motivation exists for positing a zero verbal voice prefix.

2.5 Manggarai and Rongga

Manggarai and Rongga (Arka and Kosmas 2005; Kosmas 2000) are SVO languages, showing evidence for a grammatical subject. They also show active-passive alternations without passive morphology of the kind seen in Bima. Unlike Bima, Manggarai and Rongga are highly isolating and therefore lack any verbal morphology. There are bound (clitic) forms in Manggarai, but not in Rongga. Since the same verb forms are used for

[15] The precise status of the postverbal NP cross-referencing the pronominal suffix needs further investigation. It could be that the Actor is a core argument, i.e., the NP is in an appositive or adverbial relation providing additional information to the pronominal suffix, which itself is the real core argument of the verb.

both the active and passive constructions, there is no good reason to posit a zero verbal voice prefix in these languages. The following are examples of active and passive constructions in Manggarai and Rongga:

(18) a. *Aku cero latung=k.* (Manggarai)
1s fry corn=1s
'I fry/am frying corn.'

b. *Latung hitu cero l=aku=i.*
corn that fry by=1s=3s
'The corn is (being) fried by me.'

(19) a. *Ardi pongga ana ndau.* (Rongga)
Ardi hit child that
'Ardi hit the child.'

b. *Ana ndau pongga ne Ardi.*
child that hit by Ardi
'The child was hit by Ardi.'

The (a) sentences are active, with the Actor appearing sentence-initially and the Undergoer postverbally. In the (b) sentences, the Actor appears as an Oblique, prepositionally marked as *le* (shortened to *l=*) in Manggarai and *ne* in Rongga. Evidence that the (b) sentences are syntactically passive, apart from a change in word order and the prepositional marking of the Actor, comes from other behavioural properties such as relativisation (see Arka and Kosmas 2005; Arka, Kosmas and Suparsa 2007 for details; Kosmas 2000).

2.6 Kéo and Ende

Like Rongga, Kéo (Baird 2002), Nage, and Ende are highly isolating and canonically have A-V-P structures. However, these languages exhibit no evidence for a syntactic Pivot/Subject, and neither is there any voice alternation. A notable difference between these languages and the other two Flores languages already discussed (i.e., Manggarai and Rongga) is the lack of backgrounding of the Actor when the Undergoer is fronted. That is, in Kéo and Ende, there is no structure of the type NP[P]-V-PP[A] as seen in the passive sentences of Manggarai and Rongga in (18) and (19).

However, a P argument can be fronted and is assigned pragmatic prominence. Consider:

(20) a. *Selus bhelo keli ke* (Nage-Rendhu)
Selus see mountain that
'Selus saw the mountain.'

b. *Keli ke Selus bhelo*
mountain that Selus see
'The mountain Selus saw (it).'

Sentence (20a) from the Rendhu variety of Nage exemplifies the canonical A-V-P structure in Nage. The same pattern is found in Kéo and Ende. When the P argument is fronted as in (20b), the A argument, *Selus*, remains in its preverbal position, giving rise to a P-A-V structure.

There are good reasons not to analyse the P-A-V structure in these languages as the UV. Rather, it involves topicalisation of P. Likewise, there is no good evidence for a grammatical voice in these languages of the type seen in Balinese, standard Indonesian, or Manggarai/Rongga.

Kéo and Ende, for example, appear to have no privileged syntactic unit (identified as Pivot/grammatical Subject) for certain exclusive syntactic properties such as relativisation. In Kéo, all core arguments can be relativised (Baird 2002:72), as seen in (21).[16] Crucially, when the P argument is relativised (21b-c), the A argument remains in its structural position (evidence for this is provided below). That is, the relativisation of P does not trigger a change in the grammatical status of the A argument.

(21) a. *'Imu nai tado nio [ta __ jangga déwa]* (Kéo)
3s climb unable coconut REL height tall
'He can't climb coconut (trees) that are tall.'

b. *'Ata [ta 'imu bhobha __] ké palu*
person REL 3sg hit that run
'That person that he hit ran.'

c. *Nga'o pui bili né'é pui [ta 'ata ti'i nga'o __]*
1s sweep room with broom REL person give 1s
'I swept the room with the broom that someone gave me.'

Evidence that the relativisation of P does not affect the grammatical status of A comes from the clausal negation test and the clausal adverbial insertion test.

The idea of the negation test is this: the negator (NEG) *iva* comes between A and the verb (V) in the canonical clause structure of A-V-P in Ende; i.e., A-NEG-V-P, as seen in (22a). If the A argument *ja'o* were grammatically Pivot/Subject in this structure and if there were a grammatical voice in Ende, the foregrounding of P in relativisation would be expected to be accompanied by the demotion of A from its Subject status. Such demotion is typically reflected by a change in the structural position of A. It is therefore expected that A cannot precede NEG in a relativised clause. However, this is not borne out in Ende, as seen in (22b). Thus, the A argument *ja'o* occupies the same structural position in (22a) (canonical structure) and (22b) (non-canonical structure). In Ende, there is no evidence of a change in the grammatical status of A due to the relativisation of P.

(22) a. *ja'o iva e kau* [A-NEG-V-P] (Ende)
1s NEG remember 2s
'I didn't remember you.'

---

[16] One could argue that the relativisation in (21b-c) involves not relativisation of the P, but relativisation of the TOP, which happens to be P. Hence, the bracketing for the sentence in (21b) should look like (i) or (ii):

(i) *'Ata [ta _ ['imu bhobha _] ké palu*
person REL 3sg hit that run

(ii) *'Ata [ta _ ['imu bhobha] ké palu*
person REL 3sg hit that run
'That person that he hit ran.'

While this analysis makes explicit the function of the P argument as TOP, this TOP-NP analysis does not affect the claim made in this paper; i.e., that fronting the P NP would not result in the change in the grammatical status of the A NP. Hence, there is no grammatical voice alternation in the P-A-V structure.

b. *kau* [*əta* [*ja'o iva e* __]] [P-A-NEG-V]
2s REL 1s NEG remember
'It is you who I didn't remember.'

The clausal adverbial insertion test shows a similar result. A clausal adverbial (e.g. *numai* 'yesterday') is a mobile unit in Ende, which generally appears sentence-finally as in (22a). It can also appear between the A and the verb as in (23b), but, importantly, not within the VP, that is, between the V and the P argument, as seen in (23c).

As far as the adverbial test, the idea is this: if the A argument, *Selu*, in (23a-b) were the grammatical Subject and if there were a voice alternation as in the NEG test above, we would expect that the foregrounding of P in relativisation would demote A to a non-subject status and it could no longer appear in its original canonical position. It is expected that the A argument *Selu* cannot appear before the adverbial *numai* in a relative clause. However, this expectation is again not borne out, as seen from the acceptability of sentence (23c). In other words, there is no evidence for a structural alternation (i.e., of grammatical function) as a result of P-fronting or P-relativisation in Ende.

(23) a. *Selu naka manu* ***numai*** (Ende)
Selu steal chicken yesterday
'Selu stole chicken yesterday.'

b. *Selu* ***numai*** *naka manu*
Selu yesterday steal chicken
'Selu stole chicken yesterday.'

c. **Selu naka* ***numai*** *manu*
Selu steal yesterday chicken

d. *Apa əta Selu numai naka?*
what FOC Selu yesterday steal
'What did Selu steal yesterday?'

In other Austronesian languages that have a similar clausal structure with the preverbal A NP position but show AV/UV alternation (e.g., Indonesian), the clausal negation test (and the adverbial insertion test) would force a structural position change of the A NP and a change in voice marking. Consider the following examples:

(24) a. *Amir tidak melihat gunung itu* (Indonesian)
Amir NEG AV.see mountain that
'Amir didn't see the mountain.'

b. ?**Gunung yang* [*Amir tidak melihat* __]
mountain REL Amir NEG AV.see

c. *Gunung yang* [__ *tidak Amir lihat / * melihat*]
mountain REL NEG Amir UV.see AV.see

Sentence (24a) is a negation of a canonical A-V-P structure where the negator *tidak*, like the negator in Ende, precedes the verb outside of the VP. The fronting of the P argument required in relativisation/clefting, as in (24b), is not possible if the A argument remains in its subject position, that is, before the negator; hence, (24b) is unacceptable The acceptable structure is that given in (24c) where the A NP *Amir* must follow the negator *tidak* and

precede the verb. Crucially, the verb must be in the UV (i.e., bare) form *lihat* and not the AV form *melihat*.

To conclude, there is no voice system, and therefore no zero affix, in Ende or Kéo.

3 Discussions

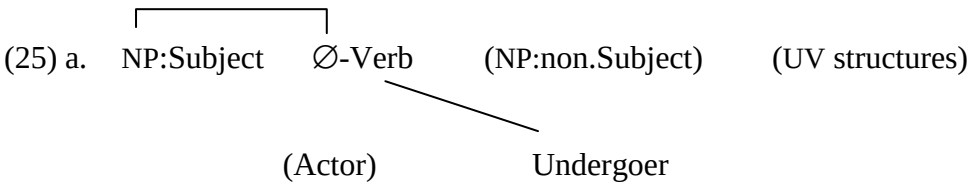
If we assume the well-accepted view that the ancient Austronesian voice systems (PAn or PMP) were multi-way voice systems similar to those currently encountered in the Philippine-type languages (see, among others, Blust 2002, 2003; Himmelmann 2002; Ross 2002; Starosta, Pawley, and Reid 1982 and the references therein; Wolff 2002), we can then say that the Balinese voice system is a simplified one, and that those of the other Austronesian languages in Nusa Tenggara discussed in this paper show further attrition, with the ultimate loss being attested by the languages of central Flores such as Ende, Nage, and Kéo. In this last section, the issue of positing a zero prefix is resumed and discussed in terms of its merit in linguistic analysis (§3.1) and linguistic theories (§3.2). A historical note is also provided in §3.3.

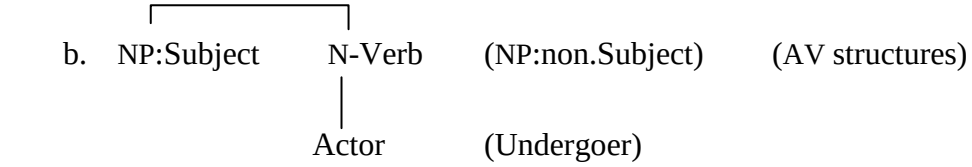
3.1 Marked or unmarked?

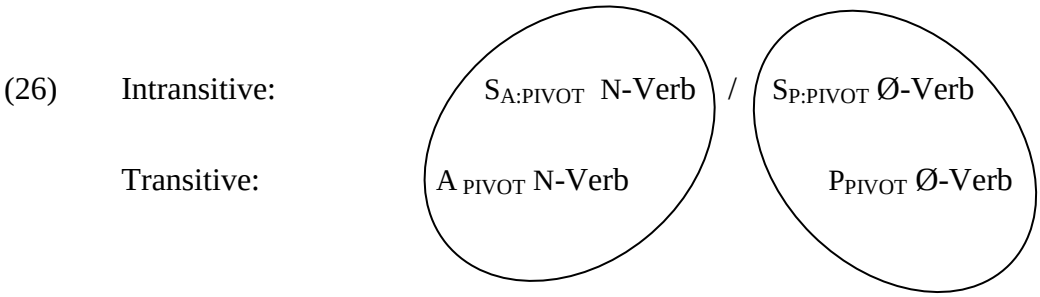
Positing a zero prefix (*Ø*-) in Balinese is well motivated, as pointed out in §1. In the analysis adopted in this paper, all verbs in Balinese are equally marked — the bare verb is marked by a zero prefix.

The basic idea of how the opposition of verbal marking works in Balinese grammar can be summarised in the diagram in (25a). The zero prefix in Balinese marks the selection of a non-Actor core argument as a Pivot or grammatical Subject. The zero prefix is a verbal coding functionally distinct from the nasal prefix (*N*-) in (25b), which marks the selection of the Actor-like argument as grammatical subject. The notation used here (i.e., the placement of the Actor(-like) argument in (25a) and the Undergoer(-like) argument in (25b) within brackets) is meant to capture the idea that the coding opposition applies to transitive and intransitive structures. That is, if there is a single core argument and if this argument is an Undergoer, then the verb is in UV form (i.e., with *Ø*-). If there are two core arguments (i.e., transitive) and if the Undergoer is selected as grammatical Subject (and the Actor remains a core argument), then the verb must also be in UV form. Having a zero prefix code UV verbs allows us to capture the basic idea in the analysis that Balinese grammar shows an active/Split-S system as far as verbal voice marking is concerned. This is represented in the diagram in (26).[17]

[17] Following common practice in work on typological linguistics, in this diagram, I use the label S to represent the sole argument of an intransitive verb and A and P to represent Actor-like and Patient-like arguments of transitive verbs. This is to highlight the idea of a Split-S system in Balinese. Note that the A and P in (26) are roughly equivalent to Actor and Undergoer in (25). It should be noted that the picture given in (26) is somewhat simplified and excludes ditransitive structures where more than one U/P argument is involved.







Let me elaborate the reasons bare verbs in other Austronesian languages of Nusa Tenggara discussed in this paper cannot be posited to have a zero voice prefix.

First, even in languages that still show a nasal prefix as in the Ngeno-Ngené dialect of Sasak, the bare verb can be used interchangeably with the nasal verbs when the actor appears in (grammatical) subject position. Therefore, unlike in Balinese, positing a zero prefix in these languages would mean that the (same) zero prefix 'marks' two quite distinct subject selections: A as subject versus U as subject. This is not an elegant linguistic analysis.

Second, given our current understanding of Sasak and other languages in Nusa Tenggara where the nasal prefix has disappeared or is disappearing, no languages other than Balinese in this area appear to have clear active/Split-S systems in their verbal morphology.

Third, languages in this area that show passivisation without verbal morphology (i.e., using bare verbs) include Sumbawa (11b), Bima (14), Manggarai (18), and Rongga (19). Given the fact that the same bare verb forms are used for structures of both active and passive voice types, the best analysis appears to be that bare verbs in these languages are unmarked, that is, without a zero prefix. As such, they are unspecified with respect to voice selection at the (morpho)lexical level. As voice alternations occur purely in syntax, making use of linear order and phrasal coding, this poses a challenge to certain linguistic theories.

3.2 Theoretical issues

Systematic voice alternations without voice morphology pose a theoretical challenge, particularly for theories that assume overt verbal markings as an indication of certain syntactic operations. The issues are relevant to both a lexical-based theory, Lexical Functional Grammar (LFG), and a movement-based theory, the Minimalist approach.

In LFG (Bresnan 2001; Dalrymple 2001; Falk 2001, among others), lexical items carry different kinds of information that constrain syntax. Passivisation and other processes that change the shape of lexical items are morpholexical operations, which change the information carried by the lexical items. The change in lexical form therefore correlates with a change in the syntactic constraints of the item. This lexical framework suits Balinese and other languages that exhibit overt affixation for syntactic alternations, such as passivisation, which changes the subcategorisation frame of a verb.

Consider the derivation processes (applicativisation and voice selection) with the transitive verb root *ambil* 'take' in (27),which shows how each derivation is registered by a particular affix.

(27) a. root: *ambil* 'take<A, U>'

b. applicativisation: *ambil-ang* 'take for <A, U: ben, U>'
take-APPL

c. voice selection

(i)	(ii)	(iii)
ng-ambil-ang	*Ø-ambil-ang*	*ka-ambil-ang*
AV-take-APPL	UV-take-APPL	PASS-take-APPL
'take.for' <A, U, U>	'take.for' <A, U:ben, U>	'take.for' <U:ben, U>
SUBJ	SUBJ	SUBJ

The transitive argument structure of the root in (27a) is represented as <A, U> (i.e., having two core arguments: an Actor and an Undergoer. The applicative *-ang* in (27b) turns the transitive verb root *ambil* into a ditransitive base, *ambil-ang* 'take for.' The applicativisation introduces a benefactive Undergoer (U:ben) into the argument-structure of the root. As noted, the ditransitive structure has an A and two U arguments. This ditransitive output, in turn, becomes an input for the voice selection process (27c). There are at least three possibilities here, depending on the selected voice types. The AV marked by the prefix *N-* (i.e., homorganic nasal) selects the A as the Pivot/Subject. The UV marked by the prefix *Ø-* selects the U:ben argument as the Pivot. The passive voice *ka-* removes or demotes the A argument from core status to Oblique, and selects the U:ben as Pivot/ Subject. In short, a verb emerges from the lexicon fully inflected, with different forms imposing different specific subcategorisation frames.

The crucial point to note for the derivation in the Balinese example is that the morphology specifies what alternation(s) have taken place and what constraint(s) the verb would impose on syntax. For example, the *ka-/-ang* affixation in *kaambilang* (27c.iii) would impose a passive structure with the promoted U:ben realised as the Pivot/Subject in the sentence, and the A would be either unexpressed or demoted to Oblique status.

However, in isolating languages like Manggarai and Rongga, there is no affix whatsoever by which morphology can tell syntax that a certain operation has taken place. It is only when the verb is used in a syntactic construction that we identify a particular voice type. That is, voice selection is achieved by means of analytic expression, for example, backgrounding the Actor and marking it with a preposition. There is a clear absence of a

morpholexical operation as observed in Balinese. Applying a morpholexical analysis such as that depicted in (27) to the isolating languages would mean applying a zero affixation to virtually all verbs for all grammatically distinct structures. This is certainly not an appealing analysis.

Applying a Chomskian transformational movement-based analysis as described in the Minimalist model (Chomsky 1995; Hornstein, Nunes, and Grohmann 2005) to the isolating languages appears to be a challenge, too. In the Minimalist framework, distinct overt morphology on the verb is a 'spell-out' of certain syntactic features. Distinct verbal voice morphology registers different kinds of 'extractions' (i.e., movement) of NPs out of the verbal nodes during syntactic derivation (Rackowski and Richards 2005). The derivation is motivated by a theory-internal case-checking mechanism. In a Minimalist analysis for the Actor voice structure in Balinese, for example, the agent NP is extracted from the verbal nodes and relevant features are spelled out on the verb as the nasal prefix. In the case of the Undergoer Voice, the patient NP is extracted and the spell-out on the verb is realised as a zero prefix (see Cole, Hermon and Yanti (2008) and Aldridge (2008) for the Minimalist analysis of Indonesian voice alternations).

In short, the crucial property of the Minimalist analysis to the Austronesian voice alternation is that distinct verbal morphology correlates with the extraction of a certain argument (Agent, Patient, etc.) from the verbal node. Again, this is fine with Balinese or Indonesian where we have distinct morphology for distinct voice types. If extended to account for voice alternation in Manggarai and Rongga, this Minimalist approach may run into the same problem as the LFG approach, precisely because of the absence of distinct morphology, assumed in this theory to be distinct spell-outs on the verb that indicate different extractions. The question is how to justify that the same (zero) spell-out correlates with two quite different extractions: one being the extraction of the Agent (equivalent to the AV structure in Balinese) and the other the extraction of the Patient in the passive. According to a Minimalist linguist (Richards 2006) in his review of Arka and Ross (2005) on this very issue in Manggarai (Arka and Kosmas 2005) and Palu'e (Donohue 2005), positing a zero prefix might not be a good idea. However, this claim is not elaborated upon any further.

Since the distinction of voice types in the Austronesian languages of Nusa Tenggara, particularly those of Flores, is encoded purely at the level of analytic expression, the data would perhaps be better couched using the insights from construction-based frameworks (Croft 2001, among others; Goldberg 1995). Lexically-based theories such as LFG could be extended to capture construction-based voice alternations, perhaps within the proposals made in Nordlinger (1998) and Ackerman (2003). Discussing the precise details of how the proposal would work with the Austronesian languages of Nusa Tenggara is beyond the scope of the present paper.

3.3 Historical notes

From a historical perspective, the discussion of zero prefix and bare verbs brings up the question of the development of voice systems and grammatical relations in the Austronesian language family. In what follows, the development of analytic passives is discussed. Due to space limitations, the discussion is focussed on the construction with third person bound pronominals that reflect PMP **ya*/**-ña*.[18]

[18] These are reconstructed forms as described in Ross (2002).

Of particular interest are constructions involving bound pronouns with backgrounded Actors. These constructions are pragmatically marked with the P being highly topical. While they resemble passives, they may not be all syntactically passives, especially in the absence of the backgrounded Actors. Consider (28), which shows the patterns seen in the languages discussed earlier and in other related languages within the so-called Proto Malayo-Sumbawan (Adelaar in press). The family tree of Malayo-Sumbawan (MS) proposed by Adelaar is shown in Figure 1.[19]

The point to be highlighted here is as follows: the constructions shown in (28) provide additional morphosyntactic evidence for the idea that Balinese, Sasak, and Sumbawa genetically form a subgroup (Adelaar in press; Esser 1938; Mbete 1990). Bali-Sasak-Sumbawa (BSS) shows a common pattern of a third person bound pronoun cross-referenced by backgrounded A (28a-e). The bound pronoun may or may not show up on the verb.

(28)					
	Balinese:		$V=(n)a_i$	(PP:agent$_i$) /*NP:agent$_i$	(a)
	Sasak:	AUX=n$_i$	V	(PP:agent$_i$) /*NP:agent$_i$	(b)
	Sumbawa:	AUX=ya $_i$	V	(PP:agent$_i$) /?*NP:agent$_i$	(c)
			i=V	(PP: agent) /?*NP:agent$_i$	(d)
			V	PP:agent /?*NP:agent$_i$	(e)
	Bima:		V-na $_i$	((P) NP:agent $_i$)	(f)
			V	PP: agent	(g)
	Manggarai/Rongga:		V	PP: agent	(h)
	Kéo/Ende/Sikka/Lamaholot:		V without backgrounded PP:agent		(i)
	Indonesian:		di-V	((P)NP:agent)	(j)
	Sundanese:		di-V	((P)NP:agent)	(h)
	Madurese:		e-V	((P)NP:agent)	(i) (Davies 2005))[20]

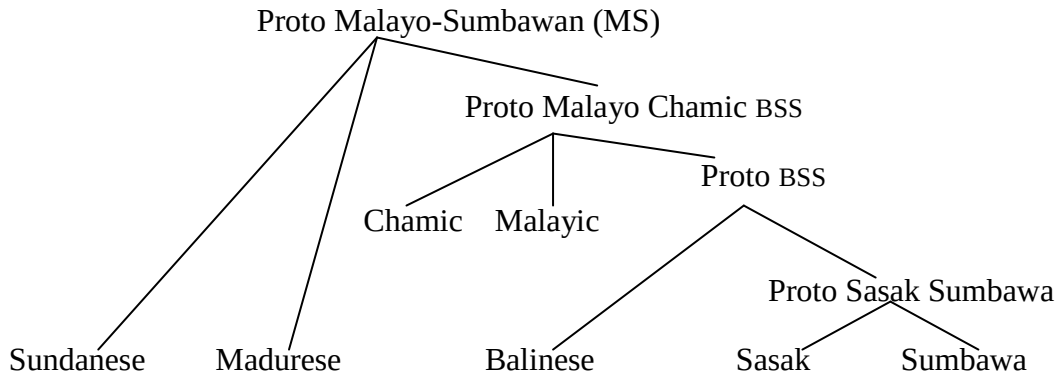


Figure 1: Family tree of Malayo-Sumbawa (Adelaar in press)

[19] I modified the direction of the branching, (i.e., putting Sundanese at the far left instead of the far right) to reflect the geography.

[20] *e-* in Madurese can appear with a non-third person Actor:

Alwi e-pokol sengko
Alwi OV-hit 1
'I hit Alwi.'

As noted, the A must be obliquely marked by a preposition in the BSS group. This sets them apart from other languages within the MS group, namely Indonesian, Madure and Sundanese. In these languages, backgrounded Actors can be NPs or PPs. In the following examples from Madurese, the Actor appears as a PP, *bi' Siti*, in (29a), and an NP, *Marlena*, in (29b).

(29) a. *Meja juwa e-saba'-i buku bi' Siti* (Madurese)
table that UV-put-E book by Siti
'Siti put the book on the table.'

b. *Bambang e-bala-agi Marlena dha' Ita*
Bambang UV-say-AGI Marlena to Ita
'Marlena talked about Bambang to Ita.' (Davies 2005, ex. 28, 22)

The realisation of the backgrounded A in BSS also sets these languages apart from Bima and the Flores languages. BSS belongs to western Malayo-Polynesian (WMP) whereas Bima and the Flores languages belong to the Bima-Sumba subgroup within central Malayo-Polynesian (CMP) (Grimes 1992; Ross 1995).[21] In Bima, the backgrounded A can appear as a PP or NP; but, unlike Indonesian/Sundanese/Madurese, Bima (together with Manggarai and Rongga) allows it with a bare verb.

Balinese is the only language within MS discussed here that has no proclitic A arguments, that is, of the type *ku*=/*kau*=/*dia*=Verb seen in Indonesian. Hence, Balinese deserves to be set apart from Sasak-Sumbawa on the one hand, and from Indonesian/ Madurese/Sundanese on the other. All A pronouns in Balinese are postverbal, and only the third person is clearly realized as a clitic, namely, *=(n)a*.[22] The third person clitic has been grammaticalised to become a passive marker in contemporary Balinese (Arka 2003; 2008a for further discussion).

(30) a. *Celeng-e adep-a teken I bapa* (Balinese)
pig-DEF sell-3.PASS by ART father
'The pig was sold by father.'

b. **Celeng-eadep-a I bapa*

c. *Celeng-e adep=a*
pig-DEF UV.sell-3
'S/he sold the pig.'

Given the family tree in Figure 1, which is established on phonological evidence, and given the patterns of backgrounded Actors with possible cross-referencing pronominals so far discussed, we can now hypothesise the development of passives, particularly analytic passives as seen in (11b) in Sumbawa.

First of all, it is reasonable to adopt the view that the PP marking requirement for the backgrounded A as encountered in BSS must be a recent innovation. That is, the backgrounded Actor argument was initially just a core argument, typically an NP. Then, it was interpreted as an oblique-like argument, which therefore licensed the PP marking. The

---

[21] The terms WMP and CMP are still used here as convenient labels of groupings of languages that have been generally considered to be related (Blust 1981, 1993; Grimes 1992). The evidence for the phylogenetic unity of these grouping labels has been questioned (see Donohue and Grimes (2008) for detailed discussion).

[22] *-na* is used when the verb is vowel-final, e.g., *aba-na* 'bring-3'.

double marking, of the type seen in contemporary Indonesian/Madurese, must have been available in the stage of Proto Malayo-Sumbawa (PMS).

The passive-like voice with cross-referencing pronominal must have developed out of the Undergoer-Voice (UV) construction (also known as PF-Patient Focus). Given the widely accepted view that PAn/PMP was relatively rich in voice morphology, UV in Proto Malayo-Sumbawan (PMS) must have been morphologically or pronominally marked. The UV coding might have been in the same prefix slot of the Actor Voice **maN-*. The system might have been like the one in Indonesian, shown in (31). At the time of PMS, there must also have already been a real passive, that is, the ancestral (proto) form of the contemporary passive *ter-* (Indonesian) and *te-* (Sasak) with the Actor being obligatorily an Oblique PP.

(31) Indonesian

Actor Voice	Undergoer Voice		
(ALL PERS)	1sA	2sA	3sA
meN-V	*ku*=V	*kau*=V	*dia*=V/*di*-V

Therefore, the bare or analytic passive of the type currently observed in Sumbawa, pattern (28e), must have been a further developmental step after the emergence of the coding of the Actor PP. That is, PP encoding might have rendered the verbal coding (e.g., *i*-(*<*ya*) in (28d) in Sumbawa) functionally redundant. The verbal marking was then dropped.

The development that led to the analytic passive construction, for example, as seen in Sumbawa, can be hypothesised to have followed the stages shown in (32a-c). Stage (a) is the one prior to PMS where a backgrounded Actor was a core argument, marked by core phrase markers. Stage (b) is the PMS stage where two kinds of marking were possible, PP and NP. Stage (c) is the PBSS stage where cross-referencing of backgrounded Actor must be in the PP. Finally, in stage (d), the bound pronominal associated with the UV disappeared, resulting in an analytic passive. At the same time, the AV marking, that is, *N-* (<**maN-*) disappeared as well (cf. §2.3).

(32) a.	Backgrounding of A:	NP:Topic	UV_i=VERB	NP:BackgroundedActor-core$_i$
b.	NP/PP A marking:	NP:Topic	UV_i =VERB	[(P) NP:BackgroundedActor(-core)]$_{PP}$
c.	PP A marking only:	NP:Topic	UV_i =VERB	[PP:BackgroundedActor(-core)]$_{PP}$
d.	Loss of UV/AV marking:	NP:Topic	VERB	[PP:BackgroundedActor]

The existence of analytic passives in Bima, Manggarai, and Rongga deserves a brief comment because these languages belong to CMP rather than WMP, but they show a passive-like structure with a backgrounded Agent PP of the type shown in (32d). Note that such an analytic passive with a backgrounded Actor, either as a PP or an NP, is not encountered in the languages of central and eastern Flores. Neither is it typically found in other Bima-Sumba/CMP languages.[23] Therefore, the (analytic) passive with a backgrounded PP in Bima, Manggarai, and Rongga must not be a property inherited from the ancestral language of Proto Flores (see Fernandes (1996) for the genealogical grouping of Flores languages), or Proto Bima-Sumba, or PCMP (if such a language ever existed; see Donohue and Grimes (2008)). The developmental stages might be different from those

[23] However, the cross-referencing of a bound A pronominal on the verb by a free NP, not by a PP, is observed in the neighbouring CMP languages of Kambera, Sumba (Klamer 1996). Klamer argues that such a structure is grammatically not passive.

hypothesised in (32). While it is reasonable to conclude that the analytic passive in Bima, Manggarai, and Rongga might not have followed the stages shown in (32), the precise history of the emergence of bare verbs, in transitive verbs in particular, or the analyticity of Flores languages in general, is far from clear.

Language contact might have complicated the picture. It has been pointed out by Adelaar (in press) that contemporary Balinese has been overwhelmingly influenced by Javanese, and this has masked the genetic link of Balinese to its neighbouring languages to the east (i.e., Sasak and Sumbawa). Balinese acquired the speech level system and lexicon from Javanese (Clynes 1989). As Lombok was occupied by the Balinese from the mid-17th to the early 20th century, the speech level system and Balinese lexicon also spread to Sasak. The islands of Sumbawa (i.e., the location of Bima) and Flores used to be under the occupation of the Goa Kingdom of South Sulawesi in the early 17th century. In addition, Flores (in particular western-central Flores where Manggarai and Rongga are now spoken) fell to the domination of the Bima Kingdom after the Kingdom of Goa was defeated with the help of the Dutch colonial force in the mid-17th century. Contact between ethnic groups through trade and political dominance in this region has been going on for centuries. However, it is still unclear whether the passive-like constructions with backgrounded Oblique Actor PPs now encountered in Bima and other languages of western Flores were acquired through contact.

4 Conclusion

The morphosyntactic status of bare verbs in the Austronesian languages of Nusa Tenggara discussed in this paper are not exactly identical. In Balinese, bare verbs can surely be analysed as having a zero prefix on the basis of formal and functional opposition with other verbs prefixed with *N-* (AV) (<PMP **maN-*), *ka-* (PASS), or *ma-* (MV). However, no good evidence exists for such an analysis in the other neighbouring languages. This is mainly due to the (gradual) disappearance of the AV prefix reflecting **maN-*. In Sasak, *N-* verbs are often used interchangeably with bare verbs in certain contexts. Bare verbs are therefore 'unmarked'; that is, they are not exclusively used for one particular voice type. In Sumbawa, the nasal prefix is only restricted to intransitive verbs and is totally lost in transitive verbs. In Bima, it is lost both in intransitive and transitive verbs.[24] In Flores languages, which are highly isolating, there is no trace of the Austronesian AV, **maN-*.

Discussion has been provided with respect to the synchronic analysis of the bare verbs. It has been pointed out that extreme analyticity that shows voice alternation is a challenge to theoretical linguistics. While it is suggested that insights from construction-based

---

[24] Bima has *ma*, which may not necessarily be prefixed to a verb. *Ma* may show constraints similar to the AV prefix *N-*. For example, it cannot be used when a transitive content question object is fronted (Arka 2009). Consider the contrast between (ii) and (iii) below:

(i) *Cou* ***ma-****mpanga janga awin?*
who ma-steal chicken yesterday
'Who stole chickens yesterday?'

(ii) *Aula Ami mpanga-na __ awin?*
what ART name steal-3s yesterday
'What did Amir steal yesterday?'

(iii) * *Au la Ami* ***ma****-mpanga-na __ awin?*

frameworks should be incorporated to allow a proper analysis of the data, the precise details of the analysis still need to be worked out.

Finally, bare verbs are the ultimate end of attrition where all verbal (voice) markings (affix and pronominal clitics) have disappeared, resulting in extremely isolating languages, as has happened to Flores languages.

Discussion has focussed on the development of bare verbs with cross-referencing backgrounded Actors, which may have led to analytic passives. It appears that historical processes of analyticity in Sumbawa, Bima, and Manggarai on the one hand, and in other languages of central and eastern Flores on the other, may not be the same. While the findings in this paper provide some support for the genealogical grouping of Bali, Sasak, and Sumbawa, the question remains of how to explain the observation that the analytic passive is encountered in Sumbawa, Bima, and languages of western Flores such as Manggarai and Rongga, but not in other isolating languages of central-eastern Flores such as Ende and Kéo. For this question, further investigation is needed to uncover the complex issues of dialect chains and contact-induced language change in the area.

References

Ackerman, Farrell. 2003. *Comparative grammar and grammatical theory : a construction-based study of morphosyntax* Stanford: CSLI.

Adelaar, K. Alexander. in press. Malayo-Sumbawan. In Peter Austin, ed. *Working papers in Bali, Lombok and Sumbawa*. London: School of Oriental and African Studies.

Aldridge, Edith. 2008. Phase-based account of extraction in Indonesian. *Lingua* 118.1440–1469.

Arka, I Wayan. 2000. On the theoretical and typological aspects of Termhood in (Eastern) Indonesian Languages (First year report, seminar paper). Jakarta: Ministry of Education and Culture.

—— 2003. *Balinese morphosyntax: a lexical-functional approach*. Canberra: Pacific Linguistics.

—— 2008a. Voice and the syntax of =a/-a verbs in Balinese. In Peter K. Austin and Simon Musgrave, eds *Voice and grammatical relations in Austronesian languages*, 70–89. Stanford: CSLI.

—— 2008b. Core argument alternations in some Austronesian languages of Nusa Tenggara, Indonesia. In Peter K. Austin and Simon Musgrave, eds *Voice and grammatical relations in Austronesian languages*, 183–227. Stanford: CSLI.

—— 2009. Attrition of Voice Morphology and Fronted Content Questions in the Austronesian languages of Nusa Tenggara, Indonesia. Paper presented 11th International Conference of Austronesian Langauges). Aussois, France.

Arka, I Wayan and Jeladu Kosmas. 2005. Passive without passive morphology? Evidence from Manggarai. In Arka and Ross, eds, 2005, 87–117.

Arka, I Wayan, Jeladu Kosmas, and I Nyoman Suparsa. 2007. *Bahasa Rongga: tatabahasa acuan ringkas*. Unpublished manuscript. Canberra: The Australian National University.

Arka, I Wayan and Malcolm D. Ross, eds. 2005. *The many faces of Austronesian voice systems: some new empirical studies*. Canberra: Pacific Linguistics.

Austin, Peter. 2001. Verbs, valence and voice in Balinese, Sasak and Sumbawan. In Peter Austin, Barry Blake and Margaret Florey, eds *Explorations in valency in Austronesian languages. La Trobe Papers in Linguistics,* vol. 11, 47–71. Melbourne: LaTrobe University.

—— 2002. Clitics in Sasak, eastern Indonesia. Paper presented at the 10th National Congress of the Indonesian Linguistics Society Conference. Denpasar, Indonesia.

Baird, Louise 2002. A grammar of Keo: an Austronesian language of East Nusantara. PhD thesis. Canberra: The Australian National University.

Blust, Robert. 1981. The Soboyo reflexes of Proto-Austronesian *S. In Robert A. Blust, ed. *Historical Linguistics in Indonesia, part 1*, 21-30. Jakarta: NUSA.

—— 1993. Central and central-eastern Malayo-Polynesian. *Oceanic Linguistics* 32(2).241–293.

—— 2002. Notes on the history of 'focus' in Austronesian languages. In Wouk and Ross, eds, 2002, n63–80.

—— 2003. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42(2).438–478.

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

Bresnan, Joan. 2001. *Lexical functional syntax*. London: Blackwell.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Mass.: The MIT Press.

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

Cole, Peter, Gabriella Hermon and Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118.1500–1553.

Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehman, ed. *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Austin Press.

Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press.

Dalrymple, Mary. 2001. *Lexical functional grammar*, *syntax and semantics*. San Diego: Academic Press.

Davies, William D. 2005. The richness of Madurese voice. In Arka and Ross, eds, 2005, 197–220.

Dixon, R.M.W. 1979. Ergativity. *Language* 55.59–138.

—— 1994. *Ergativity*. Cambridge: Cambridge University Press.

Donohue, Mark. 2005. The Palu'e passive: from pragmatic construction to grammatical device. In Arka and Ross, eds, 2005, 59–85.

Donohue, Mark and Charles E. Grimes. 2008. Yet more on the position of the languages of Estern Indonesia and East Timor. *Oceanic Linguistics* 47(1).115–159.

Esser, S.J. 1938. Talen [Languages]. In *Atlas van Tropisch Nederland*. Batavia: Koninklijk Nederlandsch Aardrijksundig Genootschap.

Falk, Yehuda N. 2001. *Lexical-functional grammar*. Stanford: CSLI.

Fernandes, Inyo Yos. 1996. *Relasi historis kekerabatan bahasa Flores*. Jakarta: Nusa Indah.

Florey, Margaret. 2001. Verbs and Valence in Alune. *LaTrobe papers in Linguistics*. 11.73-120.

Foley, William A. and Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Goldberg, Adele E. 1995. *Constructions : a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Grimes, Barbara. 1992 (12th edition). *Ethnologue*. Dallas: Summer Institute of Linguistics.

Himmelmann, Nikolaus P. 2002. Voice in western Austronesian: an update. In Wouk and Ross, eds, 2002, 7–15.

Hornstein, Norbert, Jairo Nunes and Kleanthes K Grohmann. 2005. *Understanding Minimalism*. Cambridge, UK: Cambridge University Press.

Jacq, Pascale. 1998. How many dialects are there? In Peter K. Austin, ed. *Sasak*, 67–90. Melbourne: University of Melbourne.

Jauhary, E. 2000. Pasif bahasa Bima. MA thesis. Denpasar, Indonesia: Udayana University.

Klamer, Marian. 1996. Kambera has no passive. *NUSA* 39.12–30.

—— 1998. *A grammar of Kambera*. Berlin: Mouton de Gruyter.

Kosmas, Jaladu. 2000. Argumen aktor dan pemetaannya dalam bahasa Manggarai. MA thesis. Denpasar, Indonesia: Universitas Udayana.

Kroon, Yosep B. 1998. The *isiq* construction and its grammatical relations. In Peter K. Austin, ed. *Sasak*, 105–118. Melbourne: University of Melbourne.

Mbete, Aron Meko. 1990. Rekonstruksi protobahasa Bali-Sasak-Sumbawa. PhD thesis. Jakarta: Universitas Indonesia.

Nordlinger, Rachel. 1998. *Constructive case: evidence from Australian languages*. Stanford: CSLI.

Rackowski, A. and N. Richards. 2005. Phase edge and extraction: a Tagalog case study. *Linguistic Inquiry* 36(4).565–599.

Richards, Norvin W. 2006. The many faces of Austronesian voice systems: some new emprical studies. Book Review. *Oceanic Linguistics* 45(2).501–504.

Ross, Malcolm D. 1995. Some current issues in Austronesian linguistics. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary. An introduction to Austronesian studies*, 45–120. Berlin: Mouton de Gruyter.

—— 2002. The history and transitivity of Western Austronesian voice and voice marking. In Wouk and Ross, eds, 2002, 17–62.

Shibatani, Masayoshi. 2008. Relativization in Sasak and Sumbawa, Eastern Indonesia. *Language and Linguistics* 9(4).865–916.

Shiohara, Asako. 2000. Relativization in Sumbawan. In Peter K. Austin, ed. *Sasak* 85–98. Melbourne: University of Melbourne.

Starosta, Stanley, Andrew K. Pawley, and Lawrence A. Reid. 1982. The evolution of focus in Austronesian. Paper presentated at the 3rd International Conference on Austronesian Linguistics. Bali-Indonesia.

Van Valin Jr., Robert D. and Randy J. LaPolla. 1999. *Syntax. Structure, meaning and function*. Cambridge: Cambridge University Press.

Wolff, John. 2002. Final words: the development of the focus system. In Wouk and Ross, eds, 2002, 437–449.

Wouk, Fay. 2002. Voice in the languages of Nusa Tenggara Barat. In Wouk and Ross, eds, 2002, 285–309.

Wouk, Fay and Malcolm Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

14 *Dental discrepancies and the sound of Proto Austronesian*

MARK DONOHUE

1 Dental and alveolar places[1]

Languages almost universally include oral plosives in their inventory of consonants, and of the oral consonants labial, coronal, and dorsal places of articulation can be reckoned to be the most 'basic', in terms of consistently appearing in use in different languages, and showing more contrasts than other places that are used. Coronal, in particular, covers a wide range of articulatory ground, and contrasts of place within the larger description of 'coronal' are not uncommon. Although such a contrast is rare in the languages of Europe, it is far from unheard of. The plosives of Hula, an Oceanic language of southeast New Guinea, are shown in (1), and here it is clear that the dental stops contrast with alveolar ones.

(1)	Hula	p	t̪	t	k	ʔ

Other languages utilise just the dental place, or more commonly just the alveolar place; the examples in (2) and (3) show the plosives of Eivo (Rapoisi) and Momu (Fas), two non-Austronesian languages of Papua New Guinea (from Bougainville and North-Central New Guinea, respectively).[2]

(2)	Eivo	p	t̪		k	ʔ
(3)	Momu	p		t	k	ʔ

1 I take great delight in offering this small piece of research as a footnote to some of the work that Malcolm Ross has undertaken over the years. In addition to being an inspiration for careful, accurate and topical research, Malcolm's temperament in other areas of his life, his graciousness in dealing with all that comes his way, and his generosity of spirit in all ways that I have seen, represent a goal that I would like to aspire towards, but realise that I am unlikely to achieve. Much of my own work owes not only its direction, but also much of its intelligibility, to Malcolm's patient tutoring. Much of the clarity of the work here presented is due to the insightful comments of two anonymous reviewers, and the rest follows from the editor's attention to detail. I also wish to thank David Gil for valuable discussion of the data.

2 Regrettably a great many grammars and phonological descriptions do not specify these phonetic details. In the discussion that follows I can only follow the description given by the sources I have consulted, and so probably under-represent the number of languages showing discrepancies of the sort investigated here. Equally, however, there is no reason to believe that under-representation is any more or less prevalent in any particular family, such as Austronesian, or any particular area, such as Island Southeast Asia, and so the use of statistical evaluation of different populations can be justified.

A small number of languages, such as Hawaiian (with just *p k* and *ʔ*), have no coronal stops; these cases will not be further considered here. When a language involves more than one contrastive manner of articulation with each stop series, the same three basic possibilities are found: contrastive in both dental and alveolar places, contrastive only for the dental place, or contrastive only for the alveolar place. In (4)–(6) we can see the plosive systems of Kala Kawaw Ya (Pama-Nyungan; Australia), Bilbil and Gumawana (both Austronesian, Papua New Guinea)

(4)	Kala Kawaw Ya	p	t̪	t	k
		b	d̪	d	g
(5)	Bilbil[3]	p	t̪		k
		b	d̪		g
(6)	Gumawana	p		t	k
		b		d	g

Yet another possibility is found, in which both dental and alveolar places are used, but not contrastively. Bauzi (Geelvink Bay; western New Guinea) exemplifies this pattern (note that Bauzi also displays a gap for /p/, a trait not uncommon in languages of [particularly northern] New Guinea).

(7)	Bauzi		t̪ʰ		kʰ
		b		d	g

This last is a very rare (or, possibly, under-reported) pattern globally; Maddieson (1984) lists only three languages, 1% of his worldwide survey show this pattern — one of these is Austronesian, Sundanese (Malayo-Polynesian; Java), the others being Gã from Ghana and Guahibo from Colombia. Nonetheless, this pattern is prominent in Austronesian languages, appearing in fully 3% of the Austronesian languages sampled. We can speculate on some phonetic motivations for the contrast: voiceless stops typically have a longer closure than do voiced ones, and the supralaryngeal air pressure will more easily be accommodated with a more anterior closure, thus motivating a more front point of articulation for the voiceless coronal than for the voiced one. This argument does not, however, account for the lack of reports of more palatal articulation for voiceless dorsal, compared to their voiced counterparts, and we must conclude that a purely phonetic explanation will not account for the observed frequencies. Similarly we can speculate on a possible enhanced contrast for the coronal pair, involving a non-optimal set of feature specifications (contra Clements 2003); such feature+manner bundles are not uncommon (e.g., Ladefoged and Maddieson 1996, §3 of this paper), but we still have no account for the absence of, for instance, p b t̪ d kʲ g systems, in which the dorsal stops are similarly phonetically differentiated. To what extent is the appearance of this pattern inherited, or acquired by other means?

The discrepancy, at its most constrained, can be defined as follows:

(8) A dental discrepancy can be described in a consonant system if:

---

[3] Axelson and Oliver list the coronal plosives for Bilbil/Bilibil as t̪ and d̪ but describe them in column headings as ‘LabDent’ and ‘Dent’. Malcolm Ross (pers. comm. 28 October 2008), who has collected data from the Bilbil language, does not report labiodental stops.

- there is a contrast in terms of manner (VOT, prenasalisation, laryngeal contrast) between stops in at least some places
- for coronals there is a discrepancy between contrastive stops in terms of place (within the coronal space)

We have already seen this system in Bauzi. Another plosive system that satisfies these conditions can be seen in Gã, a Niger-Congo (Kwa) language of Ghana. The plosive system of the language is shown in (9). There is a contrast, between voiceless and voiced, in all places (in most cases the voiceless series is aspirated). For the two coronal stops the contrast is not just one of VOT (here voiceless aspirated versus voiced), but also one of place, with the voiceless stop described as dental, while the voiced one is alveolar.

(9)	Gã							
		p^h	$t̪^h$		$tʃ^h$		k^h	
						$tʃ^w$		kp
		b		d	dʒ	$dʒ^w$	g	gb

I would like to introduce a further qualification in describing dental discrepancies, concerning the 'direction' of the imbalance. In (7) and (9) we have seen plosive systems in which the dental stop is voiceless, while the voiced counterpart is alveolar. In (10) we can see that in Waffa, from eastern New Guinea, it is the voiceless stop that is alveolar, while its (prenasalised) voiced counterpart is dental.

(10)	Waffa						
		p		t		k	ʔ
		mb	n̪d̪			ŋg	

Systems of the kind exemplified by Waffa are less frequent (and, as we shall see, more geographically constrained) than the more typical system with a dental discrepancy. This is described in (11).

(11) The prototypical dental discrepancy involves two coronal stops such that the stop with the greater VOT is articulated more to the front than is the other coronal stop; non-prototypical systems are found when the stop with the greater VOT has the more anterior articulation.

I shall refer to systems such as that seen in Waffa as representing a 'reversed dental discrepancy', since the position of the two contrastive stops is reversed from the more common type of dental discrepancy.

Some work suggests that the place contrast is in fact primary; in Malay, for instance, there is neutralisation for the feature [voice] in final position, with the consequence that /t̪/ and /d/ are distinguished only by place. Thus the pair of stops (in analogous environments) found in the final coda of /wujud/ 'face' and /rambut/ 'hair' contrast in terms of the place of their final segment: [wudʒʊt] versus [rɐmbʊt̪] (e.g., Abu Bakar et al. 2007). Blust (1999:325–326) similarly describes in careful detail the preservation of place contrasts between /t̪/ and /d/ when word-final devoicing occurs. The rest of this exploration addresses the question of the distribution of languages that have been reported to have a dental discrepancy, addressing the question of the evidence for reconstructing this discrepancy to Proto Austronesian.

2 Dental and alveolar places in Austronesian

Ross (1992:31) follows Haudricourt (1965:321) in reconstructing the plosive system of Proto Austronesian as shown in (12), with the sort of dental-alveolar mix for simple the coronal series that we have just examined.

(12)	Proto Austronesian	p	t̪				k	q	ʔ
		b		d	ḍ	j̪	g		
				ts					
				dz					

There are degrees of controversy associated with some of the phonetic values of these reconstructions, but none with most of them (e.g., *p* or *b*). In this comment I shall focus on the attribution of the voiceless coronal stop to a dental place while the other coronal obstruents are thought to have been alveolar, examining the grounds for reconstructing this contrast this far back in Austronesian linguistic history. I shall call this type of disparity between the voiced and voiceless members of a coronal opposition a 'dental discrepancy', as in (8).[4] Ross (1992:44) is quite explicit about the dental-alveolar discrepancy in Proto Austronesian:

> PAn *t and *d_1 were evidently the plain stops, but, as in a number of western Indo-Malaysian languages, their points of articulation were possibly different: dental [t̪] and alveolar initial [d]/non-initial [r̃].

Haudricourt reconstructed the difference in places of the coronal stops in part to explain the contrast between dental and retroflex stops in Javanese (now thought to owe at least some of its provenance to Indic influence), and to offer an account for the fact that PMP *t and *d did not merge in Proto Oceanic, while the labial and velar stops did.[5] Presaging some of the conclusions to be presented in §5 onwards, we should note that both Javanese and (it is most likely) Proto Oceanic are southern Austronesian languages.

3 Kinds of coronal mixes

Although we are concentrating here on languages involving a dental-alveolar mix for the 'same' series, it is possible for stops in the coronal space to show mixed places for a 'pair' of contrasting plosives in other ways. Table 1 presents examples of other mixes involving non-contrastive variation in the specification of the place of articulation for what is phonologically a single series of coronal stops. In all cases the paradigm of contrasts for the coronal place is split between two areas of articulation, most commonly involving the dental and alveolar places, but sometimes with one of the places being described as retroflex. We can see here that, while simple VOT contrasts are the most frequent contributor towards distinguishing the stops, prenasalisation, preglottalisation, implosion or creakiness are also possibilities. We can also see that in some cases a language distinguishes more than two coronal stops; note particularly the case of Muna, with four alveolar stops and one dental stop.

---

[4] It is, of course, quite possible for a discrepancy to appear with other stop series. In Yawa (West Papuan; western New Guinea), for instance the plosives are *p t k b d dʒ*, where *dʒ* substitutes for the expected *g* (Jones 1986). Discrepancies such as these, not involving dental/alveolar mixes, are not considered here.

[5] We note, however, that a dental discrepancy as described here is not reported with any significant frequency for Austronesian languages of the Oceanic subgroup, and is confined to two subgroups of Oceanic, suggesting that any place difference among the coronal stops was widely lost very soon after the breakup of Proto Oceanic.

Table 1: Kinds of coronal mixes

Kind of mix				Example language
dental-alveolar	t̪	d		Sundanese
(‘dental discrepancy’)	t̪ʰ	d		Gã
	t̪	ʔd		Thao
	t̪	d nd		Selayar
	t̪	nd		Kyaka Enga
	t̪ʰ t̪	nd		Huli
	t̪ʰ t̪	d nd		Palu’e
	t̪ d̪	ɗ		Tama
	t̪ʰ	t		Guahibo
(‘reversed	d̪	t		Galela
dental discrepancy’)	n̪d̪	t		Upper Asaro, Waffa
	d̪	t nt nd ɗ		Muna
	n̪d̪	tʰ	ɳʈ ɳɖ	Tinrin
dental-retroflex	t̪		ɖ	Yami, Gorum
	t̪ʰ		ʈ̪	Siona
	t̪ʰ d̪		ɖ̪	Somali
alveolar-retroflex		t	ɖ	Lelemi
				Awngi

The relevant points in a consideration of the question of dental discrepancies are:

- where do we find any dental stops in languages?
- where do we find languages with dental-alveolar contrasts?
- where do we find languages with dental discrepancies?
- to what level in the Austronesian tree should we reconstruct dental discrepancies, and to what should we attribute their appearance?

I shall address these questions in the following sections, following a discussion of the data that I shall draw on.

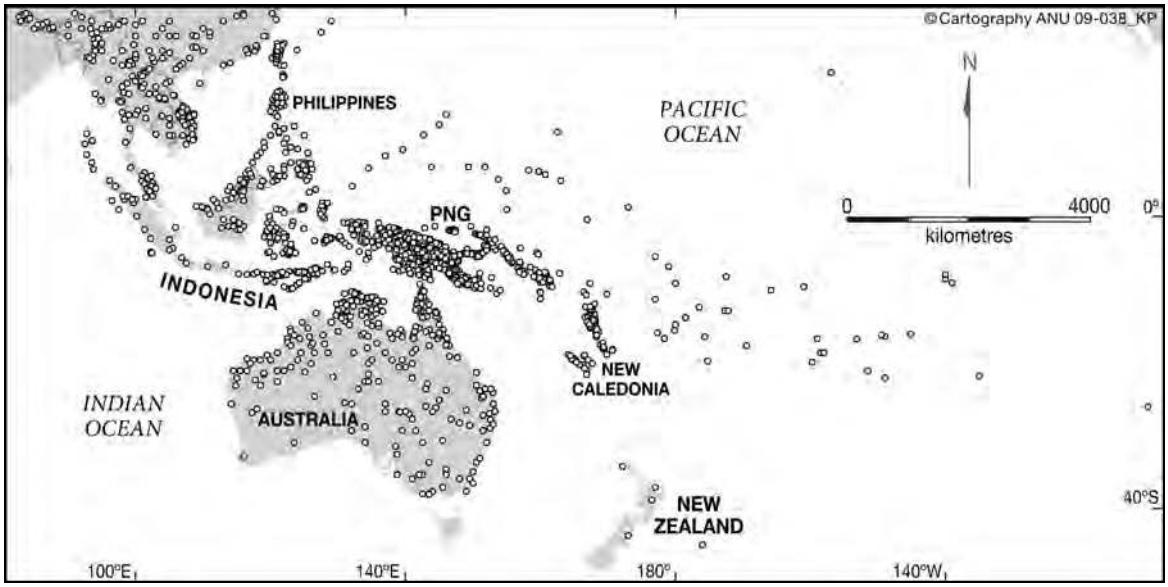
4 Data on dental discrepancies in the extended Austronesian world

I examined 1680 languages, a subset of the sample in Donohue (in preparation), approximately divided as follows:

- 680 Austronesian languages, from all areas where Austronesian languages are spoken;
- 450 ‘Papuan’ languages from most genetic groupings in and near New Guinea;

- 250 languages from Australia;
- 300 languages from mainland Asia, concentrating on Indochina but extending north to East Asia and into the Himalayas.

These languages represent a maximal census of the languages in the 'extended Austronesian' area — that region in which Austronesian languages are found, and the languages on the fringes of this region. The languages in the sample are shown in Map 1; as can be seen, coverage is quite exhaustive.



Map 1: The languages (and variant dialects) in the current sample

5 The distribution of dental stops

The different ways in which a language may contain a dental stop in its inventory, as attested in the current sample, as well as their relative frequencies, are shown in Table 2. Because of the appearance of a number of unusual patterns that are attested only in Australia, we shall consider the sample *without* the Australian languages for the purposes of determining significant skewings of the distributions.[6] From Table 3 we can see that 3% of Austronesian languages show a t̪ ~ d dental discrepancy; this is significantly higher than would be expected, based on the fact that 2% of the languages in the sample have this type of dental discrepancy.

Only 1% of languages in Maddieson's (1984) global sample have this pattern. Further, in the 1680-language sample this figure is replicated amongst the non-Austronesian languages, where we find that 1% have the dental discrepancy; compared to this, the 3% of Austronesian languages with a dental discrepancy is striking. This difference in frequency represents a statistically significant difference, much greater than can be assigned to chance ($p = 0.002$ on a two-tailed chi-squared test).

[6] Some justification for their exclusion can come from the fact that dental articulation in Australia is typically *lamino*-dental, not *apico*-dental, thus representing a different type of opposition to that discussed here, which involves apical contrasts. Furthermore, the fact that there is a complete lack of languages with a dental discrepancy in Australia might be said to justify their exclusion from the count since Australia is clearly outside the area in which dental discrepancies are found.

Table 2: Occurrences of different types of dental stops in languages in the survey

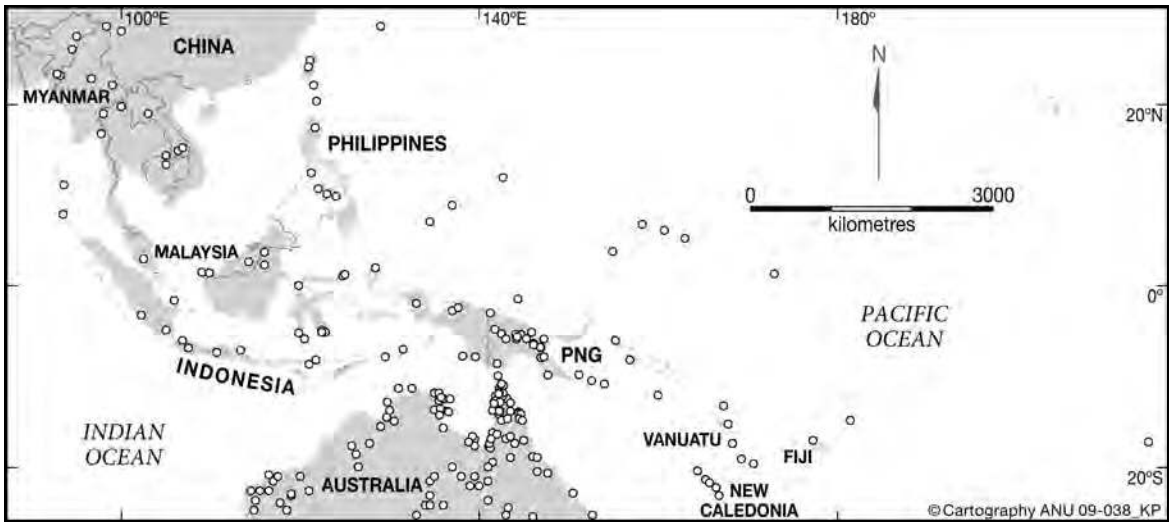
	Austronesian	non-Austronesian	Total	
Only dental stops	27	31	58	
Only alveolar stops	599	678	1277	
No coronal stops	7	5	12	[a]
Dental ≠ Alveolar	5	62	67	
Dental ≠ Retroflex	4	7	11	
Alveolar ≠ Retroflex	18	103	121	[b]
Dental ≠ Alveolar ≠ Retroflex	–	105	105	
Dental discrepancy	22	6	28	
Reversed dental discrepancy	5	7	12	
Total	683	1008	1691	

[a] Largely Australian pattern.
[b] Purely Australian pattern.

Table 3: Occurrences of different types of dental stops in languages in the survey (Australia excluded)

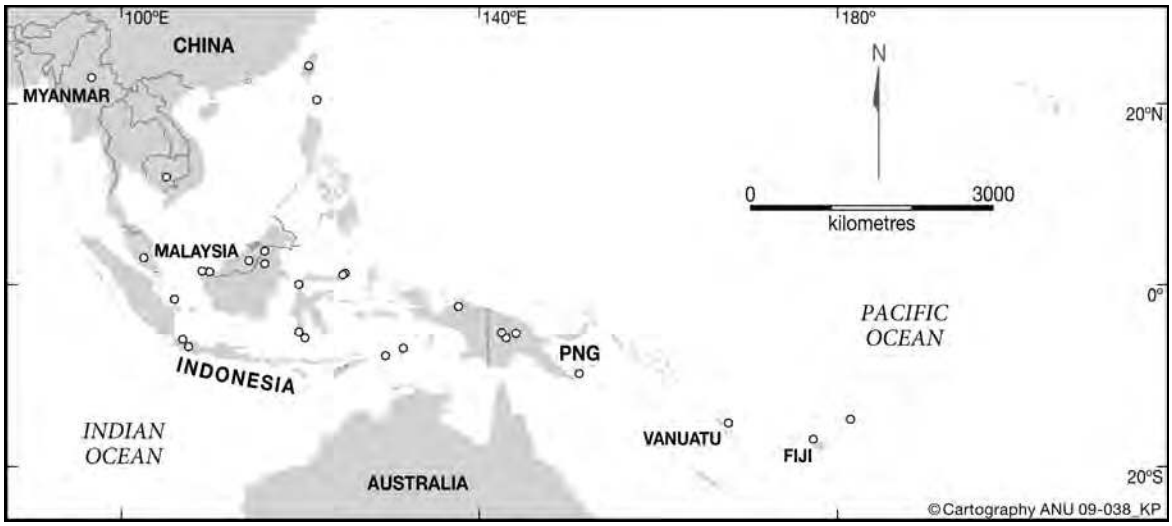
	Austronesian	non-Austronesian	Total	p (χ^2)
Only dental stops	27	29	56	0.92
Only alveolar stops	599	642	1241	0.67
No coronal stops	7	5	12	0.51
Dental ≠ Alveolar	5	5	10	0.94
Dental ≠ Retroflex	4	6	11	0.54
Alveolar ≠ Retroflex	18	21	39	0.73
Dental discrepancy	**22**	**6**	**28**	**0.002**
Reversed dental discrepancy	5	5	10	0.94
Total	683	719	1407	

Examining the distribution in space of languages with dental stops is not particularly enlightening. As can be seen in Map 2, they are found across most of the extended Austronesian area and beyond, being significantly absent only from mainland East Asia and parts of Australia (predominantly in the east, but also including substantial areas in the western desert), and also appearing sporadically throughout most of New Guinea. Most of these dental stops represent the sole coronal stop (or stop series) found in the language in question, and so are do not indicate a contrast, phonemic or phonetic, with another coronal stop. While it is more common for a sole coronal series to be alveolar, rather than dental, it is not so uncommon to find dental articulation in a language.



Map 2: Languages with any dental stops

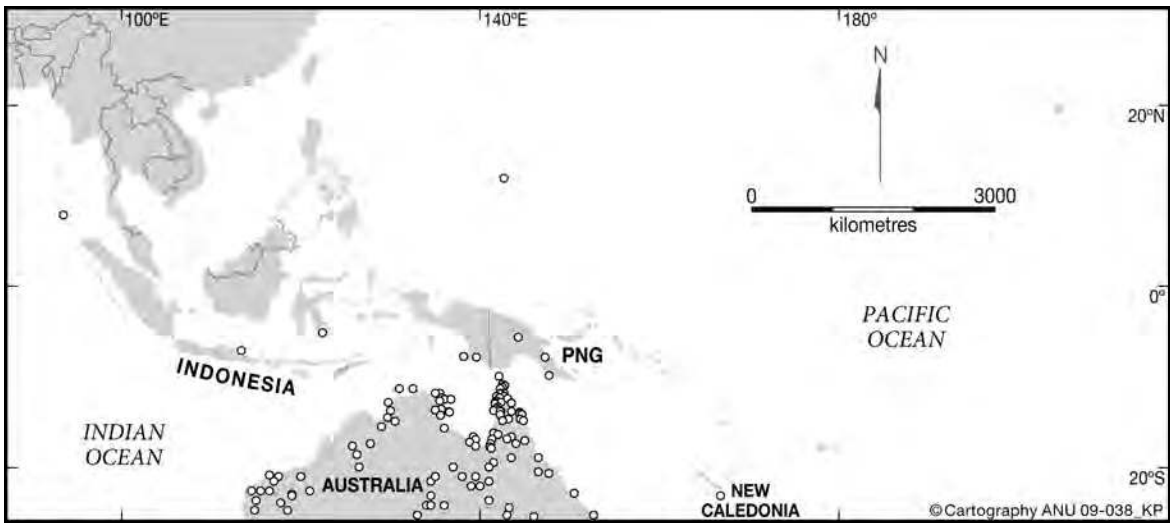
The distribution in space of languages with a dental discrepancy is much more revealing. Map 3 shows the languages with the t̪ ~ d dental discrepancy pattern, and it is immediately apparent that this pattern is not evenly distributed across the Austronesian world. While there is a more-or-less continuous, though scattered, realisation of this contrast in the languages of Indonesia and New Guinea, reports of this contrast are lacking for languages in the Philippines (where some languages, such as Cebuano, are reported with dental stops, but with both members of the coronal voice opposition showing dental articulation) and in remote Island Melanesia, where less languages are reported as showing any use of dental articulation and almost none with a dental discrepancy.[7]



Map 3: Languages with a dental discrepancy (t̪ ~ d)

7 One reviewer notes that '[t]he picture for the Philippines may be largely a result of under-reporting', and mentions Aklanon as having an unreported (in the published literature) dental discrepancy. It is quite likely that such lack of attention to detail underlies many phonetic descriptions, not just of Austronesian languages and not just those of the Philippines, as described in footnote 2.

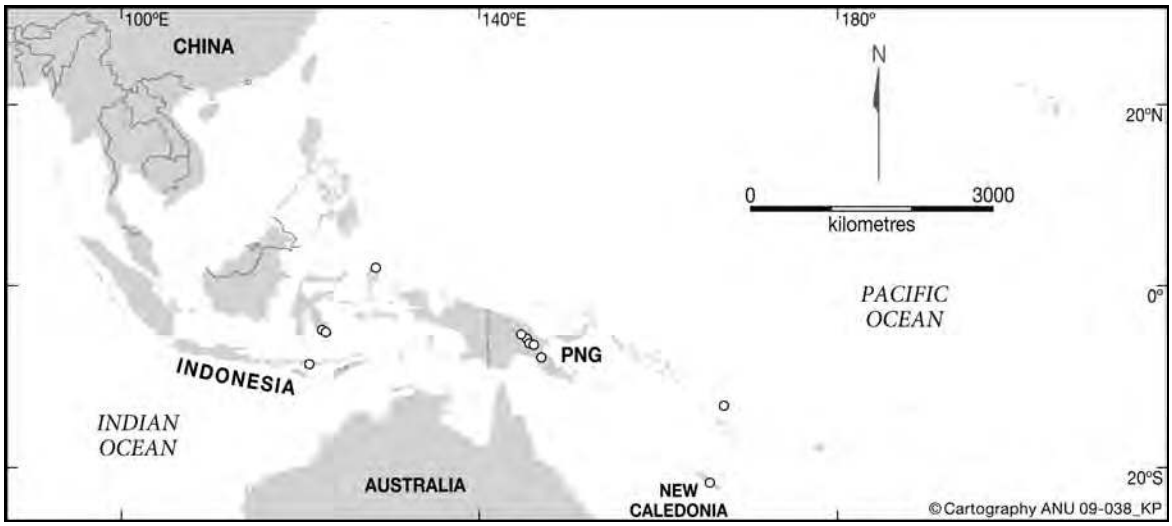
The occurrence of languages with a dental series that contrasts with another coronal series is shown in Map 4. The trend is quite clearly for languages with dental contrasts to cluster closer to Australia, especially in New Guinea but also as far as Madurese (on Java) and Muna (in Southeast Sulawesi). Only Nancowry Nicobarese in the west, and Grand Couli from New Caledonia in the east go against this generalisation.[8] Significantly, the area with contrastive (lamino-)dental stops is immediately south of the area in which dental discrepancies are found, suggesting strongly that dental discrepancies can be thought of, in some sense, as being a 'fringe' version of a dental contrast (see the discussion of Blust 1999 and Abu Bakar et al. 2007 in §2).



Map 4: Languages with phonologically contrastive dental stops

The reversed dental discrepancy is also found to the north of the dental contrast area, but in less languages and scattered across a more constrained dispersal. The region in which reversed dental discrepancies are found is almost perfectly described as the limits of the spread and influence of Papuan languages; Donohue (2007a) describes the possible classification of Tambora, found just west of the westernmost diamond in Map 5, as 'Papuan', and some of the arguments for considering a Papuan substrate in remote Melanesia, Vanuatu and New Caledonia, are presented in Blust (2005, 2008) and Donohue and Denham (2008a). The significance of this more restricted distribution, compared to the more common 'normal' dental discrepancy seen in Map 3, and their appearance with respect to the dental contrast languages shown in Map 4, will be discussed in the following section.

8 Dental discrepancies of the dental-retroflex pattern are also found immediately west of the map in the Austro-Asiatic Munda languages Gorum and Gutob, and to a lesser extent in Gtaʔ, and possibly others in South Asia. This suggests that the pattern in Khmer is an ancient Austro-Asiatic one, and not recently acquired through contact with others.



Map 5: Languages with a reversed dental discrepancy (d̪ ~ t)

6 The distribution of dental discrepancies

The data in the maps are open to various interpretations, but one that appears probable, and which takes account of the different facts presented as well as other knowledge about the likely prehistory of the region, involves the following observations:

1. The 'homeland', in the Pacific area, for languages with phonemic contrasts between dental and alveolar stops is Australia. We can identify a fringe of languages nearby in southern New Guinea showing similar contrasts, and only a small number of others away from Australia.
2. Dental discrepancies primarily occur immediately to the north of Australia in New Guinea and Indonesia, with only minor appearances to the west, east and north.
3. Reversed dental discrepancies are even more restricted to being found close to Australia (plus isolated outliers in northern Vanuatu and New Caledonia) than are normal dental discrepancies

On this basis, we would, on geographic grounds alone, make the following suppositions to attempt to explain the distribution of the dental discrepancy pattern:

1. The appearance of dental contrasts and reversed dental discrepancies in just the areas for which we must suppose evidence of a significant pre-Austronesian linguistic presence implies that Australia and its northern environs are a centre for dental discrepancies.
2. While *reversed* dental discrepancies are found in both Austronesian and non-Austronesian languages with equal frequency, languages with normal dental discrepancy are much more frequent in Austronesian languages. This implies that, if they are not original to the family but have been acquired through contact, the distribution and type of dental discrepancies in Austronesian reflects a 'founder effect'. Austronesian languages, in other words, display only a subset of the diversity that is associated with dental stops.

3. Within the Austronesian languages large areas, mainly in the Philippines and Solomons, are found in which dental discrepancies are not reported. It is thought that these areas represent recent expansions that effectively wiped out potential earlier diversity (e.g., Blust 1991, 2000), possibly including an earlier dental discrepancy. Perhaps more significantly, only three languages in the northern part of the Austronesian area are reported with a dental discrepancy (Pazeh, Thao and Southern Ivatan), and only two that are not Malayo-Polynesian (see Figure 1).
4. A parsimonious explanation of the presence and distribution of dental discrepancies in Austronesian languages should suppose that, rather than being an inherited feature (and thus worthy of reconstruction to Proto Austronesian), it is an acquired one in the Malayo-Polynesian languages.

An obvious problem with the preceding account is the presence of a dental discrepancy in the two northern languages. Southern Ivatan can be explained: although it is found very far to the north in the Malayo-Polynesian area, there is no linguistic evidence that allows us to assume that Southern Ivatan is any 'higher' in the Austronesian tree than any other Malayo-Polynesian language (Ross 2005), just as there is no linguistic evidence for the Malayo-Polynesian spread across island southeast Asia to have proceeded north-to-south (Donohue and Denham to appear; I grant, of course, that it proceeded from Taiwan to the islands to the south, but the nature of the dispersal within the islands is unresolved). The same contact (or, more likely, language shift) scenario that resulted in the acquisition of dental discrepancies can also be invoked to account for the dental discrepancy in Southern Ivatan. We know that we must posit contact from the south through the Ivatan region because of the presence of Proto Austronesian reconstructions for items (*CebuS 'sugarcane', and *manuk/*qayam 'chicken') that are known not to be original to Taiwan, but to have been transported there (presumably prior to the breakup of Proto Austronesian) (Donohue and Denham to appear), and see §7.

Pazeh and Thao, in central Taiwan, are not so simply dismissed. Figure 1 shows the Austronesian family tree, following Donohue and Grimes (2008) on the removal of the Central-Eastern Malayo-Polynesian and Central Malayo-Polynesian nodes, with the relative positions of the dental discrepancy languages shown in the abstract with dots; Pazeh and Thao are the only first-order branches of Austronesian with a dental discrepancy.

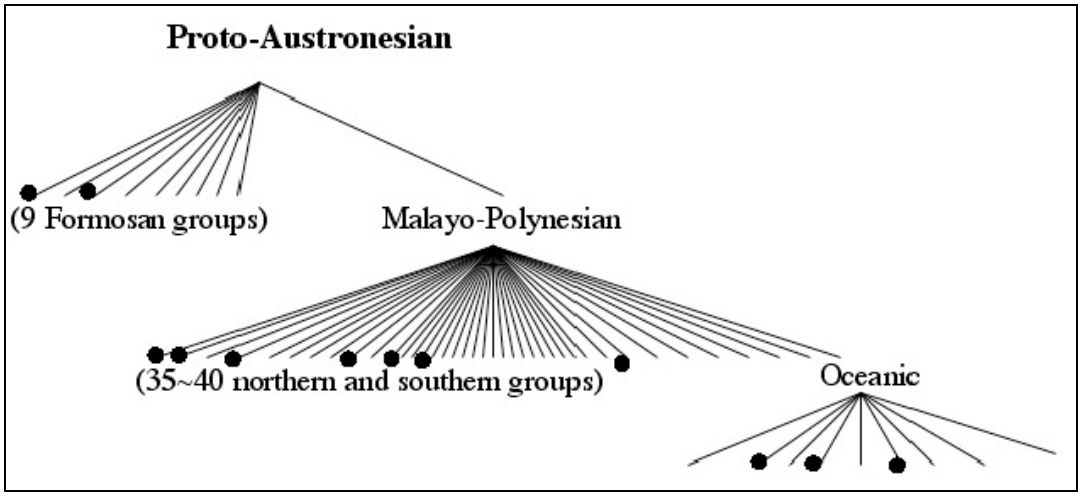


Figure 1: The Austronesian family tree

7 Interpretation

On basic principles we should, as per Ross (1992), reconstruct the dental discrepancy to Proto Austronesian; it occurs in two widely separated primary subgroups of Austronesian, Thao (or Pazeh) and Malayo-Polynesian, and so should be reconstructed. But is this the whole story?

There are just two witnesses beyond Malayo-Polynesian, Pazeh and Thao, and certainly no evidence that they have ever been in the area in which non-Austronesian languages elaborate the use of the dental place of articulation. But that does not necessarily mean that the area in which non-Austronesian languages elaborate the use of the dental place of articulation cannot have come to central Taiwan.

While Taiwan is home to the Austronesian languages, not all that is Austronesian finds its home in Taiwan. It is well-known that sugarcane, for example, is indigenous to New Guinea (Grivet et al. 2004); yet sugarcane is one of only three food crops that can be reconstructed to Proto Austronesian (as *CebuS). If the plant term can be reconstructed to Proto Austronesian, its appearance in the protolanguage must be presumed to predate the dispersal of Malayo-Polynesian languages across Island Southeast Asia. On the other hand, if the plant itself cannot be assumed to be indigenous to Taiwan, it must be presumed to have been brought to Taiwan before the Malayo-Polynesian dispersal in order to be incorporated in the lexicon of Proto Austronesian. This accords with our developing understanding of the role of maritime interaction and trade prior to the Austronesian dispersal (e.g., Torrence and Swadling 2008; Donohue and Denham 2008b; Denham and Donohue 2009; Denham, Donohue and Booth 2009), which suggests that prior to the dispersal of the Malayo-Polynesian languages in Southeast Asia there was already a robust mosaic of trading cultures, transporting obsidian and food crops about the archipelago, and in some cases into the Pacific.

As noted in §6, the distribution of dental discrepancies is essentially found (erratically) in Indo-Malaysia and New Guinea (Map 3), forming a fringe north of the area in which dental stops contrast with other coronals (Map 4). Reversed dental discrepancies are even more geographically restricted, found only in the fringe area that evidences a Papuan presence (Map 5). This geographic clustering is unlikely to have arisen by chance, and strongly suggests that the existence of stops in two coronal places, whether or not they form a phonemic contrast, is areally predicted. The fact that Australia is the centre of this area clearly reflects facts about the human population prior to the arrival of Austronesians; and if this is true, then the appearance of this pattern in the Austronesian languages of the area similarly reflects the acquisition of this pattern as a substrate by the languages of the most recent polity to arrive in the area, that associated with the Austronesians. Numerous other linguistic facts attest to the presence of robust substratal influence in this area (e.g. Donohue 2007b; Donohue and Denham to appear), and the fact that there is no known or proposed subgrouping that contains the languages with a dental discrepancy means that we must invoke some kind of geographically-driven explanation.

If we assume that dental discrepancies (and dental contrasts) are not an originally Austronesian feature, then we must explain how a dental discrepancy came to be found in Pazeh and Thao, and why the reverse discrepancy is not much less frequent in Austronesian. The second point can simply be attributed to a founder-effect; while dental discrepancies (of either direction) might have been (relatively) common in pre-Austronesian Indo-Malaysia, this feature did not spread into Austronesian ‘evenly’, but through initial exposure to (socially influential) languages with ‘normal’ dental

discrepancies. The appearance of a dental discrepancy in Pazeh and Thao is, on first impression, not easily ascribed to external factors: both languages have been spoken in the middle of Taiwan, not close to any coasts (at least during the 'ethnographic present'). On the other hand the modern demography of Austronesian Taiwan bears little resemblance to what was first encountered 400 years ago, let alone 4000. It is not implausible to suggest that a linguistic precursor to one of these (currently geographically close) languages acquired a dental discrepancy from an external trading language, spoken by members of an influential and prestigious social group, and that this pronunciation spread in a limited area. It might be that we do not need to reconstruct a dental discrepancy to Proto Austronesian after all; or even to assume that the Thao and Pazeh cases represent independent innovations.

Appendix: data and sources

Languages with a ('normal') dental discrepancy.

Non-Austronesian languages: Bauzi (Geelvink Bay, Indonesia; Briley 1976), Duna (Duna-Bogaya, New Guinea; San Roque pers. comm.), Gorum (Anderson and Rau 2008), Gtaʔ (DeArmond 1976, Griffiths 2008), Gutob (DeArmond 1976, Anderson 2008), Huli (Engan, Papua New Guinea; Rule 1977), Khmer (Austro-Asiatic, Cambodia; John Ohala via David Gil, pers. comm. 9 April 2009), Kyaka Enga (Engan, New Guinea; SIL-PNG), Palaung (Austro-Asiatic, Burma; Shorto 1960), Semelai (Austro-Asiatic, Malaysia; Kruspe 2004).

Austronesian languages: Ambae (Austronesian, Vanuatu; Hyslop 2001), Bauan Fijian (Austronesian, Fiji; Geraghty 1995a), Duau (Austronesian, Papua New Guinea; SIL-PNG), Fordata (Austronesian, Indonesia (east); Marshall and Marshall 1992), Indonesian (Austronesian, Indonesia (west); own notes), Ivatan (Southern) (Austronesian, Philippines (Northern); Heye and Hidalgo 1967), Kayan Kenyah (Austronesian, Indonesia (Borneo); Sorriente 2006), Kayan, Uma Juman (Austronesian, Indonesia (Borneo); Blust 1977), Konjo (Austronesian, Indonesia (central); Friberg 1995), Lom (Austronesian, Indonesia (west); Nothofer 1994), Ngaju Dayak (Austronesian, Indonesia (west); Brunelle and Riehl 2002), Pendau (Austronesian, Indonesia (central); Quick 2008), Putoh (Òma lóngh) (Austronesian, Indonesia (Borneo); Sorriente 2006), Selako (Austronesian, Malaysia (Borneo); Adelaar 2005), Selayar (Austronesian, Indonesia (central); Mithun and Basri 1986), Sundanese (Austronesian, Indonesia; Clynes 1995), Thao (Austronesian, Taiwan; Li 1978, Blust 2003), Tondano (Austronesian, Indonesia (central); Sneddon 1975), Tonsawang (Austronesian, Indonesia (central); Sneddon 1978), Western Fijian (Austronesian, Fiji; Geraghty 1995b), Wetan (Austronesian, Indonesia (east); De Josselin de Jong 1987).

Other languages cited.

Awngi (Afro-Asiatic, Ethiopia; Hetzron 1969), Bilbil (Austronesian, Papua New Guinea; SIL-PNG), Eivo (Rapoisi) (South Bougainville, Papua New Guinea; SIL-PNG), Gã (Kwa, Niger-Congo, Ghana; Ablorh-Odjidja 1968), Galela (West Papuan, Indonesia; Wada 1980), Guahibo (Guahiban, Colombia; Kondo and Kondo 1967), Gumawana (Austronesian, Papua New Guinea; SIL-PNG), Hawaiian (Austronesian, USA; Elbert and Pukui 1979), Hula (Austronesian, Papua New Guinea; Short n.d.), Kala Kawaw Ya (Pama-Nyungan, Australia;

own notes), Kyaka Enga (Engan, Papua New Guinea; SIL-PNG), Lelemi (Kwa, Niger-Congo, Ghana; Hoftmann 1971), Momu (Kwomtari, Papua New Guinea; SIL-PNG), Muna (Austronesian, Indonesia; van den Berg 1989), Palu'e (Austronesian, Indonesia; own fieldnotes), Siona (Tucanoan, Colombia; Wheeler and Wheeler 1962), Somali (Afro-Asiatic, Somalia; Armstrong 1964), Tama (Nilo-Saharan, Chad; Tucker and Bryan 1966), Tinrin (Austronesian, New Caledonia; Osumi 1995), Upper Asaro (Gorokan, Papua New Guinea; SIL-PNG), Yami (Austronesian, Taiwan; West 1995).

References

Ablorh-Odjidja, J.R. 1968. *Ga for beginners*. Accra: Waterville Publishing House.

Abu Bakar, Husni, Karthik Durvasula, Nadya Pincus and Timothy McKinnon. 2007. Incomplete laryngeal neutralization in Standard Malay. Paper presented at the 11th International Symposium on Malay/Indonesian Linguistics (ISMIL 11), Manokwari, Indonesia.

Adelaar, K. Alexander. 2005. *Salako or Dameà. Sketch grammar, texts and lexicon of a Kanayatn dialect in West Borneo*. Wiesbaden: Otto Harrowitz.

Anderson, Gregory D.S. 2008. Gtaʔ. In Anderson, ed. 2008, 682–763.

Anderson, Gregory D.S. and Felix Rau. 2008. Gorum. In Anderson, ed. 2008, 381–433.

Anderson, Gregory D.S., ed. 2008. *The Munda languages*. London: Routledge.

Armstrong, L.E. 1964. The phonetic structure of Somali. *Mitteilungen des Seminars fur Orientalische Sprachen (Berlin)* 37(3).116–161.

Blust, Robert A. 1977. Sketches of the morphology and phonology of Bornean languages 1: Uma Juman (Kayan). In *Papers in Bornean and western Austronesian languages*, 8–122. Canberra: Pacific Linguistics.

—— 1991. The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30(2).73–129.

—— 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35.272–294.

—— 1999. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2).321–365.

—— 2000. Chamorro historical phonology. *Oceanic Linguistics* 39(1).83–122.

—— 2003. *Thao dictionary*. Taipei: Academia Sinica.

—— 2005. Book review: The Oceanic languages. *Oceanic Linguistics* 44(2).544–558.

—— 2008. Remote Melanesia: one history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* 47(2).445–459.

Briley, David. 1976. Bauzi phonology. *Irian* 5(1).47–65.

Brunelle, Marc and Anastasia Riehl. 2002. Vowel laxing in Ngaju Dayak. *Working Papers of the Cornell Phonetics Laboratory* 14.1–15.

Clements, G.N. 2003. Feature economy in sound systems. *Phonology* 20.287–333.

Clynes, Adrian. 1995. Sundanese. In Tryon, ed., 1995, 459–468.

DeArmond, Richard C. 1976. Proto-Gutob-Remo-Gtaq stressed monosyllabic vowels and initial consonants. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta, eds *Austroasiatic studies, part I*, 213–227. Oceanic Linguistics Special Publication No.13. Honolulu: University of Hawai'i Press.

De Josselin de Jong, J.P.B. 1987. *Wetan fieldnotes*. Dordrecht: Foris.

Denham, Tim and Mark Donohue. 2009. Pre-Austronesian dispersal of banana cultivars west from New Guinea: linguistic relics from eastern Indonesia. *Archaeology in Oceania* 44.18–28.

Denham, Tim, Mark Donohue, and Sara Booth. 2009. Horticultural experimentation in northern Australia revisited. *Antiquity* 83(321).634–648.

Donohue, Mark. 2007a. The Papuan language of Tambora. *Oceanic linguistics* 46(2).520–537.

—— 2007b. Word order in Austronesian: from north to south and west to east. *Linguistic Typology* 11(2).351–393.

—— in preparation. *Patterns of Pacific sounds*.

Donohue, Mark and Tim Denham. 2008a. The language of Lapita: Vanuatu and an early Papuan presence in the Pacific. *Oceanic Linguistics* 47(2).365–376.

—— 2008b. Banana (Musa Spp.) domestication in the Asia-Pacific region: linguistic and archaeobotanical perspectives. *Ethnobotany Research and Applications* 7.293–332.

—— to appear. Farming and language in island southeast Asia: reframing Austronesian history. *Current Anthropology* 51(2).

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.

Elbert, Samuel H. and Mary Kawena Pukui. 1979. *Hawaiian grammar*. Honolulu: University of Hawai'i Press.

Friberg, Timothy. 1995. Konjo. In Tryon, ed., 1995, 563–572.

Geraghty, Paul. 1995a. Eastern Fijian (Bauan). In Tryon, ed., 1995, 919–923.

—— 1995b. Western Fijian (Nadrongā). In Tryon, ed., 1995, 925–930.

Griffiths, Arlo. 2008. Gutob. In Anderson, ed., 2008, 633–681.

Grivet, L., C. Daniels, J.C. Glaszman and A. D'Hont. 2004. A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research and Applications* 2.9–17.

Haudricourt, André G. 1965. Problems of Austronesian comparative philology. *Lingua* 14.315–329.

Henderson, Eugenie J.A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15, 400–434.

Hetzron, R. 1969. *The verbal system of Southern Agaw*. Berkeley/Los Angeles: University of California Press.

Heye, J. and C.A. Hidalgo. 1967. An outline of Southern Ivatan phonology. *General Linguistics* 7(2).105–120.

Hoftmann, H. 1971. *The structure of the Lelemi language*. Leipzig: VEB Verlag Enzyklopädie.

Hyslop, Catriona. 2001. *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. Canberra: Pacific Linguistics.

Jones, Linda K. 1986. Yawa phonology. In *Papers in New Guinea Linguistics* No. 25, 1–30. Canberra: Pacific Linguistics.

Kondo, V. and R. Kondo. 1967. Guahibo phonemes. *Phonemic systems of Colombian languages,* 89–98. Oklahoma: Summer Institute of Linguistics/University of Oklahoma, Norman.

Kropp Dakubu, Mary Esther. 1987. *The Dangme language: an introductory survey*. Basingstoke: Macmillan/Accra-North: Unimax Publishers.

Kruspe, Nicole. 2004. *A grammar of Semelai*. Cambridge: Cambridge University Press.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

Li, Paul Jen-kuei. 1978. The case-marking systems of the four less-known Formosan languages. In S.A. Wurm and Lois Carrington, eds *Proceedings of the Second International Conference on Austronesian Linguistics, fascile 1*, 569–615. Canberra: Pacific Linguistics.

Maddison, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Marshall, Craig and Sarah Marshall 1992. Reduplication in Fordata. In Donald A. Burquest and Wyn D. Laidig, eds *Descriptive studies in languages of Maluku*, 23–30. *NUSA: Linguistics studies of Indonesian and other languages in Indonesia 34*. Jakarta: Universitas Katolik Atma Jaya.

Mithun, Marianne and Hasan Basri. 1986. The phonology of Selayarese. *Oceanic Linguistics* 25, 210–254.

Nothofer, Bernd. 1994. Bahasa Lom – dari mana asalnya? In Liberty P. Sihombing, Multamia R.M.T. Lauder, L. Pamela Kawira, Nitrasattri Handayani, eds *Bahasawan Cendekia: seuntai karangan untuk Anton M. Moeliono*, 193–208. Jakarta: P.T. Intermasa [Fakultas Sastera Universitas Indonesia].

Osumi, Midori. 1995. *Tinrin grammar*. Oceanic Linguistics Special Publication No. 25. Honolulu: University of Hawaii Press.

Quick, Phil. 2008. *A grammar of the Pendau language*. Canberra: Pacific Linguistics.

Ross, Malcolm. 1992. The sound of Proto Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31(1).23–64.

—— 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2).1–24.

Rule, W.M. 1977. *A comparative study of the Fee, Huli and Pole languages of Papua New Guinea*. Oceania Linguistics Monographs 20. Sydney: University of Sydney.

Short, Lillian Mary Theakston. 1963. The phonology and grammar of the Hula language with vocabulary and translation and notes of other dialects of the Hood Bay District. MA thesis. Adelaide: University of Adelaide.

Shorto, H.L. 1960. Word and syllable patterns in Palaung. *Bulletin of the School of Oriental and Asian Studies* 23, 544–567.

SIL-PNG. various publications, mostly from the *Organised Phonology Data* series (which most frequently appear without clear author affiliation or dates) [Available online at http://www.sil.org/pacific/png/]

Sneddon, James N. 1975. *Tondano phonology and grammar*. Canberra: Pacific Linguistics.

—— 1978. *Proto-Minahasan: phonology, morphology and wordlist*. Canberra: Pacific Linguistics.

Sorriente, Antonia. 2006. *Mencalèny and Usung Bayung Marang: a collection of Kenyah stories in the Òma lóngh and Lebu' Kulit languages*. Jakarta: Atma Jaya University Press.

Torrence, Robin and Pamela Swadling. 2008. Social networks and the spread of Lapita. *Antiquity* 82, 600–616.

Tryon, Darrell, ed. 1995. *Comparative Austronesian dictionary*: Berlin: Mouton de Gruyter

Tucker, A.N. and M.A. Bryan. 1966. *Linguistic analyses: the non-Bantu languages of north-eastern Africa*. Oxford: Oxford University Press for International African Institute, London.

van den Berg, René. 1989. *A grammar of Muna*. Dordrecht: Foris.

Wada, Yuiti. 1980. Correspondence of consonants in North Halmahera languages and the conservation of archaic sounds in Galela. In Naomichi Ishige, ed. *The Galela of Halmahera. A preliminary survey*, 497–529. Osaka: National Museum of Ethnology.

West, Anne. 1995. Yami. In Tryon, ed., 1995, 315–320.

Wheeler, A. and M. Wheeler. 1962. Siona phonemics. *Studies in Ecuadorian Indian languages 1*, 96–111. Oklahoma: Summer Institute of Linguistics/University of Oklahoma, Norman.

15 *Deictic to relativiser in Oceanic*

ROBERT EARLY

1 Introduction

This paper looks at the relationship in some Oceanic languages between particles which are designated in traditional part of speech classifications as 'relative pronouns', but which are called here 'relativisers', and those which occur as deictics.

A relativiser is the frequently obligatory head of a relative clause, which is an embedded or subordinated clausal component which modifies a nominal head of the matrix or higher level clause. This nominal may take a range of semantic roles and grammatical functions in the matrix clause, as well as having a range of semantic roles and grammatical functions within the embedded relative clause, of which it is also an argument. The grammatical functions of the nominal in relation to the relative clause embedded under it are determined by the Noun Phrase Accessibility Hierarchy (Keenan and Comrie 1977). English relativisers are forms like *that*, *which*, *who*, and *whose*, in contexts like:

(1) [*The bread* [***which*** *you bought yesterday*]$_{\text{RELCL}}$]$_{\text{NP}}$ *is already finished*.

Deictics are those particles which occur in languages with the function of identifying the spatio-temporal location or orientation of arguments, or the spatio-temporal relationship between arguments, including tracking referents within the flow of discourse. These forms often appear in two-way oppositions like English *this – that*, *here – there*, in contexts like:

(2) *Put* ***this*** *one* ***here*** *and* ***that*** *one near* ***that*** *other one over* ***there***.

Three-way oppositions are frequently found in Oceanic languages, e.g. Māori constructions involving *nei* 'near speaker', *nā* 'near hearer, proximate', or *rā* 'distant'.

The relationship between relativisers and deictics is of interest because in many languages, the usually single form which occurs as relativiser is homophonic with one of the several forms which occur with deictic function. Certainly it is not unusual in Oceanic languages for this kind of situation to occur, and in their typological overview in *The Oceanic Languages* (TOL), Lynch, Ross and Crowley stated (2002:53) that

> … subordinate markers often perform other functions in these languages. Relative clause markers, for example, are often similar or identical in shape to demonstratives, and reason clauses are often expressed by means of a causal preposition.

A particular example of multifunctionality of grammatical particles has been described for Austronesian Saaroa (Taiwan) by Radetzky (2004). There is a form *kana'a* 'that', a distal demonstrative with some other functions, and *ka*, a marker of reason clauses, but both also 'mark definite noun phrases, participate in genitive constructions, and act as the head of relative clauses' (Radetzky 2004:426).

However, where these situations obtain, interesting questions arise about the status and provenance of these forms. In the particular case here, can the homophony be regarded as simply coincidental, or is it polysemy, with the relativiser and deictic considered to be the 'same' multi-functional particle? If there is some relationship between the two particles, or the two functions of the one particle, what does that have to say about these respective functions? And if the two functions are expressed by the same form, can one be understood to be primary, and the other a secondary development or functional extension of it?

This paper will conclude that Oceanic languages display a typologically common grammaticalisation path going from deictic to relativiser, and that this development has happened several times independently in various subgroups of Oceanic.

2 Relativiser and deictic in Lewo, Bislama, and English

It has already been noted that deictic-relativiser homophony is common, but not knowing this as a junior fieldworker, it actually seemed strange and exotic to me on first encounter with the Lewo language (Epi island, Vanuatu), that the same form *naɲa* was used to indicate the seemingly markedly distinct functions of relativiser and deictic.

In Lewo, there is a pair of regular demonstrative particles, which occur in short and long forms, *ni/nini* 'this; proximal' and *ne/nene* 'that; distal', which allow participants to be located more specifically (3).

LEWO

(3) *Yeririna la* ***nene*** *a-pimi* *nakoneva.*
people PL DEM 3PL.S-R.come yesterday
'Those people came yesterday.'

However, there are two other particles whose deictic function is restricted to the context of participant identification in discourse. In particular, they are used to bring back into the foreground of the discourse any previously-mentioned participant. These anaphoric particles are *nena,* and *naɲa*, and it is the latter of these that we are interested in here. Example (4) shows this particle functioning as a discourse deictic a number of times, and (5) shows it twice as a relativiser, where the embedded clause in the first instance is a verbless nominal clause.

LEWO

(4) *Ana yermare* ***naɲa****-ne* *Ø-mẽke* *meta* *kunus* ***naɲa****,*
CONJ devil DEIC-DISTAL 3SG.S-R.stay up hill DEIC

Ø-kira *Ø-pito* *Ø-mãl* *ninsis* ***naɲa****.*
3SGS-glance 3SGS-R.go.down 3SGS-R.see child DEIC
'Then *that* (previously-mentioned) devil there stayed up on *that* (previously-mentioned) hill, looked down (and) saw *that* (previously-mentioned) child.'

(5) *Yermarua Esau,* ***naña*** *wo-na apua Lakawe, Ø-puyu*
old.man Esau REL spouse-3SG.POSS grandparent Lakawe 3SG.S-R.climb

Ø-pa Ø-te mrakinita na kesen ***naña*** *me-se e-a.*
3SG.S-R.go 3SG.S-sit rafter GEN kitchen REL 1PLEXCL.S-stay PREP-3SG
'Old man Esau, *who* is the husband of grandmother Lakawe, climbed up and sat on the rafter of the kitchen in *which* we were staying.'

The apparent strangeness of this homophony of the deictic and the relativiser was heightened by the observation that Bislama, Vanuatu's English-lexifier variety of Melanesian Pidgin English, which often reflects the structural features of the vernacular languages, employs two quite different morphemes for these two distinct functions. The only demonstrative particle in Bislama is *ya*, which highlights reference to a participant without the identification of any particular locational or directional specification, such that it will be contextually determined as translatable by English *this* or *that*:

BISLAMA

(6) *Pikinini* ***ya*** *i stap ron.*
child DEIC PRED.MKR CTS run
'This/that child is running.'

For relativisation, Bislama has a distinct relativiser *we*:

BISLAMA

(7) *Pikinini* ***we*** *i sik tumas hem i gel blong hem.*
child REL PRED.MKR sick very 3SG PRED.MKR girl GEN 3SG
'The child who/that is really sick is his daughter.'

It was looking at the translations of sentences like the pair above which finally brought about the rather delayed realisation that the dual functionality of particles like *naña* was actually a feature of my own mother tongue as well! The English form *that* can be found as a portmanteau 'distant singular' member of the demonstrative paradigm *this*, *that*, *these*, *those* (8), and also as one of a number of morphemes (*who*, *whose*, *which*, *that*) used to relativise embedded clauses as modifiers of nominal constituents (9).

(8) ***That*** *child is running.*
[DEM N]$_{NP}$ [ ]$_{VP}$

(9) *The child* ***that*** *is really sick is his daughter.*
[DEF N [REL VP]$_{RELCL}$]$_{NP}$ [V NP]

The multifunctionality of Lewo *naña* no longer seemed so exotic, but the question of how this situation might arise, and which of the functions might be secondary developments, remained.

3 Relativisers and deictics in Oceanic

A query along these lines to Malcolm Ross, whom I was fortunate to have as a PhD supervisor at the time, resulted in him making a comment which was quoted in my dissertation:

> Malcolm Ross (pers. comm.) considers that evidence from other Western Oceanic languages indicates that historically, the deictic function of 'relative pronouns' like *naña* is probably the original one, and that the function of introducing relative clauses was a subsequent extension of this (Early 1994:430, fn.5)

It intrigued me how anyone could know this, and despite this comment, my misunderstanding of the direction of change continued into the sketch of Lamen in Lynch, Ross and Crowley (2002), where I stated that '[a]nother anaphoric particle is formed from the relative pronoun plus the proximate demonstrative, *naña-ni*' (Early 2002:674). Now, much later, it is a great joy to explore this further in a volume of papers honouring Malcolm for the impact and import of his life and work.

A first question then is to what extent do Oceanic languages employ forms which are homophonous to deictic particles as relativisers? The layout and format of the 43 grammar sketches found in Lynch, Ross and Crowley (2002) provide a very accessible database for exploring this. The intention of that volume was to provide information about previously not well-known Oceanic languages, but there was also a deliberate attempt to provide a selection that is representative of the 'major genetic or geographical groupings' (2002:x) of Oceanic. Each of the sketches contains a section on 'Articles and demonstratives' (§2.3), and another on 'Relative clauses' (§2.8), allowing for a ready comparison to be made of the forms that occur with these functions in each language.

The use of a deictic particle is clearly not the only relativisation strategy employed in the languages surveyed, and this is also the case more widely. Nikolaeva (2005:503) notes that the four basic strategies employed in the world's languages for relative clause formation are 'nonreduction, pronoun retention, relative pronouns, and gapping'. Note these are not exclusive types, as, for example, gapping can occur with or without relative pronouns. The construction types which result can be indicated by the following English sentences, mostly ungrammatical for the purposes of demonstration, where the relativised subject of the main clause ('the child') is the object of the embedded clause:

nonreduction:	[She brought the child] has malaria.
	[Which child she brought (him)] has malaria.
pronoun retention:	[The child (that/whom) she brought him] has malaria.
relative pronoun:	[The child whom she brought (him)] has malaria.
gapping:	[The child she brought] has malaria.

With regard to relative pronouns, it is also noted that

> the modified noun is represented in the RC by means of special pronouns, which are normally related to demonstratives and/or interrogatives (Nikolaeva 2005:504)

This statement reinforces the relatedness of relativisers and deictics, but says nothing about the direction of dependency or origin.

With regard to Oceanic, Lynch, Ross and Crowley are of the view that '[t]here was probably no explicit relative clause marker' (i.e. relative pronoun or relative particle) in Proto Oceanic (2002:89). This is something of an argument from silence, with no body of evidence from which any 'POc relative clause marker is reconstructible' (p.80). Furthermore, Ross (1988:233) states that '[i]n most WM [western Melanesian] Oceanic SVO languages, a relative clause is structurally identical to a dependent clause and simply follows the noun phrase it modifies'.

The second strategy mentioned above, pronoun retention, involves the occurrence of a 'trace' or 'resumptive' pronoun retaining or marking the role of the extracted head. This is considered to have been a feature of POc: 'A trace of the relativised noun phrase remained in the relative clause' (Lynch, Ross and Crowley 2002:80), and as well POc relative clauses might have involved either of the other strategies mentioned, nonreduction and gapping. The lack of an overt marker would certainly have left open the possibility of one developing in the daughter languages, and from the many possible scenarios, Lynch, Ross and Crowley (2002:80) narrow them down to four main relative clause marking strategies found in Oceanic languages. They are noted here, but in fact do not fully account for all the languages which are described in *The Oceanic Languages*. The number of languages which make use of each strategy is noted:

i. no marking [28]
ii. a relative clause marker or subordinating morpheme before the clause [20]
iii. a demonstrative before the clause [13]
iv. a demonstrative at the end of the relative clause [5].

The possibility of (iii) and (iv) above occurring in POc is not discounted, but any such use would have been 'pragmatically determined and not grammaticised', as there was 'no obligatory morphological marking of a relative clause' (Lynch, Ross and Crowley 2002:80). Note that there is no mention in this listing of pronoun retention, or the occurrence of what are sometimes called trace elements in relativised clauses, which reflect the grammatical relation of the extracted nominal which is the head of the relative clause. This is, in fact, common in Oceanic languages, as in the Lewo example (10), but is not regarded as a feature which specifically indicates subordination or relativisation.

LEWO

(10) *Sine-na* *poni* *kasi* *naɲa* *ne-paeme* *e-a.*
gut-3SG.POSS 'forget' year REL 1SG.S-REAL.be.born PREP-3SG
'She has forgotten the year in which I was born/that I was born in (it).'

While the four strategies listed above from Lynch, Ross and Crowley do predominate across the 43 languages in *The Oceanic Languages*, a more detailed inspection shows up an additional five strategies, each of which are only found in one or two languages:

v. the head noun takes a third person possessor suffix [1]
vi. a non-deictic relative marker occurs at the end of the relative clause [2]
vii. the relative clause is both preceded by a demonstrative and followed by some other particle [2]
viii. a relative marker or subordinating particle occurs, as well as an obligatory demonstrative enclitic on the subordinated verb [2]
ix. the relative clause is followed by a clitic marking the number of the head noun, and an optional demonstrative [1].

Further, it is not the case that each language follows only one of these nine distinct strategies, but many languages appear to utilise two or three of them, under conditions that are not always stated.

The following two tables list each of the languages whose sketches are found in Lynch, Ross and Crowley (2002). Table 1 notes the relativisation strategies that are employed for each language, and Table 2 indicates for each language whether or not there is any formal resemblance or apparent relationship of the relativiser with any kind of deictic particle. The first or second order subgroup of Oceanic to which the language belongs is also given, following the designations in Lynch, Ross and Crowley (2002):

ASM	Admiralties and St Matthias
WO	Western Oceanic
CEO	Central Eastern Oceanic
SJ	Sarmi-Jayapura
NNG	North New Guinea
PT	Papuan Tip
MM	Meso Melanesian
SES	South East Solomonic
UV	Utupua-Vanikoro
SO	Southern Oceanic
CP	Central Pacific
MC	Micronesian[1]

In *The Oceanic Languages* the sketches are arranged in a basically west to east ordering, and that is followed here, meaning that languages from the same subgroup are not always grouped together.

Recall that:	strategy i	is no marking
	strategy ii	is a preposed non-deictic relativiser
	strategy iii	is a preposed deictic/demonstrative as relativiser
	strategy iv	is a postposed deictic/demonstrative as relativiser
	strategy v	is third person marking
	strategy vi	is a postposed non-deictic relativiser
	strategy vii	is a preposed deictic, and postposed other particle
	strategy viii	is a relativiser plus demonstrative enclitic
	strategy ix	is a pronominal number clitic, and optional demonstrative

[1] The authors or compilers of each of the sketches are clearly identified on pp.vii–viii of *The Oceanic Languages*. They are: Mike Anderson, Simon Corston-Oliver, Ross Clark, Terry Crowley, Lucille S. D'Jernes, Mark Donohue, Robert Early, Paul Geraghty, Deborah Hill, Rex Horoi, Dorothy Jauncey, John Lynch, Catherine McGuckin, Bill Palmer, Malcolm Ross, Hans Schmidt, Joyce Sterner, and Darrell Tryon. Ross, Lynch and Crowley made multiple contributions.

Table 1: Relativisation strategies

	Language	Group	Strategy								
			i	**ii**	**iii**	**iv**	**v**	**vi**	**vii**	**viii**	**ix**
1	Kele	ASM	✔		✔						
2	Mussau	ASM	✔	✔	✔						
3	Sobei	SJ	✔	✔	✔						
4	Tobati	SJ	✔								
5	Kairiru	NNG	✔		✔						
6	Takia	NNG	✔						✔		
7	Arop-Lokep	NNG		✔							
8	Jabêm	NNG			✔	✔					
9	Gapapaiwa	PT	✔								
10	Sudest	PT	✔			✔					
11	‘Ala’ala	PT	✔				✔	✔			
12	Bali-Vitu	MM	✔			✔					
13	Kaulong	NNG	✔						✔		
14	Siar	MM	✔			✔					
15	Taiof	MM	✔								
16	Banoni	MM	✔	✔							
17	Sisiqa	MM						✔			
18	Roviana	MM		✔							
19	Kokota	MM	✔							✔	
20	Gela	SES	✔								
21	Longgu	SES									✔
22	Arosi	SES	✔			✔					
23	Buma	UV		✔							
24	Mwotlap	SO		✔							
25	Sakao	SO	✔	✔							
26	Tamabo	SO	✔	✔							
27	Raga	SO		✔							
28	Vinmavis	SO								✔	
29	Port Sandwich	SO	✔								
30	Southeast Ambrym	SO		✔							
31	Lamen	SO		✔	✔						
32	Ifira-Mele	CP	✔		✔						
33	Sye	SO		✔							
34	Anejom̃	SO	✔		✔						
35	Cèmuhî	SO		✔	✔						
36	Xârâcùù	SO	✔		✔						
37	Iaai	SO	✔	✔							
38	Ulithian	MC		✔							
39	Puluwatese	MC		✔	✔						
40	Rotuman	CP	✔	✔							
41	Nadrogā	CP	✔		✔						
42	Niuafo’ou	CP	✔	✔	✔						
43	Marquesan	CP	✔	✔							

The forms employed by each language are listed in Table 2. Languages using strategy i obviously have no form to be listed, while for those which use strategies ii or vi, the non-deictic relativiser is given. For those languages which employ strategies where the same form is used as both deictic and relativiser (strategies iii, iv, vii, viii, and ix), Table 2 gives that form, along with other relevant notes. The glosses given for each form are those used in the original source sketch grammars. The reverse arrow ← in the Notes column means that the form used for the strategy listed is given in the previous column.

Table 2: Forms employed as relativisers

	Language	Group	Forms	Strategy notes
1	Kele	ASM	*eti* 'anaphoric particle'	usually i, sometimes iii ←
2	Mussau	ASM	*ateba* 'post modifying quantifier, singular specific'	usually i, sometimes ii or iii ←
3	Sobei	SJ	*=mau* 'new topic enclitic'	i but also ii/iii ←
4	Tobati	SJ		i
5	Kairiru	NNG	*nai* 'demonstrative'	i but in text[2] can find iii ←
6	Takia	NNG	*e*, *a*, *o*: demonstratives	i but often vii:← plus termination of definite enclitic *=n* or boundary marker *ak/=k*
7	Arop-Lokep	NNG	*yo* 'REL'	optional ii ←
8	Jabêm	NNG	*tɔʔ*, *taŋ*, *te*, *nɔʔ*, *naŋ*, *ne*: short forms of deictics	iii ← and/or iv
9	Gapapaiwa	PT		i
10	Sudest	PT	*=ma* 'that (previously mentioned or known)'	i or iv ←
11	'Ala'ala	PT	*kau-na* 'person-3sg'	i, v, vi: v: head noun takes third person possessor suffix *-na* 'sg' or *-ta* 'pl'; vi ← postposed relative marker
12	Bali-Vitu	MM	*=ini* 'that'	i or iv ←
13	Kaulong	NNG	*tin* 'DEM'	i or vii: preceded by ← and followed by *men* 'delimiter'
14	Siar	MM	*na* 'proximate post-modifer', *ning* 'intermediate post-modifier', or *ning na*	i (for S of RC) or iv ←

[2] 'A relative clause is … unmarked' in Kairiru (Ross 2002:209), but *nai* is a non-selective demonstrative meaning 'near neither speaker nor hearer' (p.206), and this form seems to function as a relativiser (p.210):

moin	*nai*	*kyau*	*u-ɾ̃im*	*ceik*	*e-rib*
woman	that	1sg	1sg-see:3sg	stringbag	3sg-carry

'The woman I saw was carrying a stringbag.'

	Language	Group	Forms	Strategy notes
15	Taiof	MM		i
16	Banoni	MM	*kam* 'REL'	i or ii ←
17	Sisiqa	MM	*ne*: phrase final definite article	vi: ← at end of RC
18	Roviana	MM	*sapu* 'REL'	ii: ←
19	Kokota	MM	demonstrative enclitic on verb	i or viii: *ta* 'SUB' + ←
20	Gela	SES		i
21	Longgu	SES	*=na*	ix: followed by clitic marking number of head noun, and optional DEM ←, reduced form of *nina* 'that'
22	Arosi	SES	*(a)ni, (a)si, (a)na*	i or iv: ←
23	Buma	UV	*po* 'REL'	ii: ← or *pe* 'compl'
24	Mwotlap	SO	*a* 'REL'	ii: ←
25	Sakao	SO	*ɾm* 'DEM that'	i or ii: but only example of i has a form ← following head noun; re ii: *θ-* 'identificatory nominal prefix' on verb following relativised subject
26	Tamabo	SO	*mwede* 'REL'	i or ii:, ← glossed as 'particular.one' and not listed as a demonstrative
27	Raga	SO	*be* 'REL'	ii: ←
28	Vinmavis	SO	*nen* 'REL' *ŋe* or *nen* 'DEM'	viii: ← following head noun + ← at end of RC
29	Port Sandwich	SO		i
30	Southeast Ambrym	SO	*xa* 'REL'	ii: ←
31	Lamen	SO	*na(p̃a)* 'REL' or *na* 'DEM'	ii or iii: ←
32	Ifira-Mele	CP	*gani* 'REL'	i or iii: ← but also have *gani* 'LOC.PRO' as oblique anaphor and demonstrative
33	Sye	SO	*mori* 'REL'	ii ←
34	Anejom̃	SO	*aan* 'he'	i, or iii: ← pronoun which could become REL?

	Language	Group	Forms	Strategy notes
35	Cèmuhî	SO	ii: *naa* 'general' iii: *cɛ* 'proximate', *ne* 'intermediate', or *naa* 'distant, invisible'	ii ← or *cɛli* 'INDEF', *li* 'DEF' iii ←
36	Xârâcùù	SO	*bwâ* 'distant and invisible'	i or iii: relativised NP takes demonstrative like ← or *a* 'near speaker', *nä* 'near addressee', *nöö* 'distant from both speaker and addressee', *mwîrî* 'anaphoric'
37	Iaai	SO	*a* 'REL'	i or ii ←
38	Ulithian	MC	*læ* 'REL', *we* 'REL', also demonstratives *lee* 'proximate', *laa* 'intermediate, *wee* 'distant, not visible'	ii ←
39	Puluwatese	MC	relativisers: *yeray* 'anyone who', *ne* 'who, which', *ye* 'the one who/which'	ii ←, or iii when relativised NP has DEM, REL is omitted
40	Rotuman	CP	*ne* 'REL'	i or ii ←
41	Nadrogā	CP	*mānī* 'anaphoric demonstrative'	i or iii: ← being reanalysed as a relative clause marker
42	Niuafo'ou	CP	*'eni*, *'ena*, *ē*: demonstratives	i, ii, iii: ii: preverbal tense particles or nominal marker *ko*, or *ke* 'subjunctive' iii appositional particle *'a* followed by ← 'the one which'
43	Marquesan	CP	*ai* 'anaphoric trace'	i or ii: ii: *te* 'DEF' for head of RC as subject, possessive *ta* for head as object, otherwise ←

These data clearly show that not only is there a great deal of variation in the strategies that are employed (Table 1), but also, for languages which employ the same or similar strategies, there is also a great deal of variation in the forms which occur as relativisers, or as deictics or as other particles with a relativising function (Table 2). Simple inspection down the listing shows the following forms glossed as 'REL': *yo*, *kam*, *sapu*, *po*, *a*, *mwede*, *be*, *nen*, *xa*, *naña*, *gani*, *mori*, *a*, *læ*, *we*, *yeray*, *ne*, *ye*, *ne* — certainly a rather unlikely cognate set. Historical connection is evident with the number of deictic forms containing partials like *-ni* and *-na* or suchlike, but in each language, the formal similarity of the deictic and the relativiser suggests that the duality of function has occured within that language, rather than deictic and relativiser forms being inherited separately from an earlier parent language.

Despite this variation across Oceanic, we need to consider if languages that belong to lower level subgroups within Oceanic might not display some similar features in how they form their relative clauses, and whether or not deictic particles occur with this additional function.

The table below uses the subgroup designations from *The Oceanic Languages*. This is not an attempt to engage in historical reconstruction by numbers, and readers of this volume will be those who know that relative frequencies of similar forms in daughter languages cannot compete with the Comparative Method in order to carefully reconstruct language history. However, what the table does show is that not only is there a wide range of forms appearing in relative clause constructions, and not only is there a large number of relative clause formation strategies, but it is very difficult to find a correlation between these relative clause formation strategies and membership among the higher order subgroups of Oceanic.

Table 3: Relativising strategies in Oceanic languages

1st order subgroup	AD	SM	WO					CEO						Total
2nd order subgroup			SJ	NG	PT	MM	Total	SES	UV	SO	CP	MC	Total	
No. of languages	1	1	2	5	4	6	17	3	1	13	5	2	24	43
No. of strategies	2	3	3	5	4	5		3	1	4	3	2		
i	1	1	2	3	4	4	13	2		6	5		11	28
ii		1	1	1		2	4		1	9	3	2	15	20
iii	1	1	1	3			4			4	3	1	8	14
iv				1	2	1	4	1				1	1	5
v					1		1							1
vi					1	1	2							1
vii				1			1							1
viii						1	1			1		1	1	2
ix								1				1	1	1

Note that where a language uses more than one strategy, both are counted, so there are cases where the number of strategies employed exceeds the number of languages.

The most frequent strategy overall (strategy i, no marking at all) occurs in 65% of cases, but in nearly all languages like this, one or more other strategies are also used. In fact, there are only five languages in the sample which use strategy i only, and no other. The next most frequent strategy overall is non-deictic relativisers (strategy ii) occurring in 47% of languages, and deictics as relativisers (strategy iii + iv) in 44%. In other words, around half of the languages will have a genuine relativiser, and somewhat less than half of languages will employ a relativiser that is clearly relatable to an existing deictic particle. It needs to be kept in mind however, that while some forms are considered to be 'genuine relativisers', they could also have derived from earlier forms such as deictics which are no longer current. Keeping in mind that, as noted earlier, comparativists have not been able to

reconstruct any definitive relativiser for POc, and noting that strategy i is also the strategy that is found in more subgroups than any other, it is more likely that the preponderance of strategy i represents a retention of the original lack of a relative pronoun, rather than being the result of successive losses of an original form in various subgroups. The majority of languages in the sample still preserve this original strategy, but as noted, only five of them do so exclusively. This suggests that for whatever reasons, languages without any overt relative clause formation strategy must be subject to pressure to develop alternative strategies to reflect this syntactic operation.

Looking at the two larger first order subgroups, Western Oceanic and Central Eastern Oceanic, three broad observations can be made:

- WO languages are more likely to use strategy i, no marking, than any other (in 77% of languages)
- CEO languages are more likely to use strategy ii, non-deictic relativiser, 62.5%, but strategies i and iii are also common.
- The less common strategies (iv-ix) occur more frequently in WO languages (53% of languages) than in CEO languages (26% of languages), suggesting that there is greater diversity and complexity of relative clause formation strategies in the Western languages compared to the Central-Eastern ones.

Nikolaeva (2006:504) says that the use of an overt relativiser is 'typical of most European languages but quite infrequent crosslinguistically', but the data shows this (strategy ii) occurring in nearly half of the Oceanic languages, and in the case of CEO, in well over half.

Looking now at the second order subgroups, and discounting those with just one or two languages, it can be noted that there are cases where strategy i is used in every language in the sample (e.g. in all four Papuan Tip and all five Central Pacific languages, admittedly a small proportion of the totals of such languages), but there are no cases where any of the other strategies are used in all languages in a subgroup. This reinforces the picture that many languages: (i) have preserved the presumed original pattern of no marking of relative clauses; but (ii) have subsequently and independently developed alternative strategies for this function.

Indeed, the notable degree of variation in relative clause formation strategy, and in the forms of deictics and other particles employed with this secondary function, is significant. We are only able to talk of the Oceanic subgroup of Austronesian because phonological and morphosyntactic features have been inherited down the family tree, and in many areas, this inheritance is pervasive and regular, but there is little evidence of this in regard to the development of relativisers and relative clause formation strategies at the levels of relatedness obtained across the languages in the sample.

It might be expected that a greater level of similarity based on shared inheritance would be found at lower-level localised groupings, and an example of this is the languages of Epi, Vanuatu. The Lewo form *naɲa* discussed above (examples [4] and [5]) is related to the same form *naɲa* for Lamen in the table, and both are related to Baki *nabano* [na^{m}bano]: all three of these Epi languages form a small grouping within Central Vanuatu, and the formal similarity noted here is paralleled by these anaphoric particles also functioning as relativisers in each language.

However, at the level of the parent languages of such closely related localised subgroups or linkages, or earlier, such similarity is not found. The diversity of forms and structures across and within the higher-order subgroups of Oceanic is indicative of independent development, or parallel innovation at lower stages, similar to that claimed for the development of verb stem initial consonant alternations in different Vanuatu and other Oceanic languages (Crowley 1991).

It is interesting to wonder why there are some elements of structure (e.g. pronouns, negative markers) which careful comparative work has shown can be clearly traced through various stages of Oceanic language history, while there are others, as in the case of relativisers, where this does not seem possible. Without supporting evidence, any claims about how or why this happens can only be speculative, but perhaps it says something about the stability of features with varying salience or complexity. This stability would affect how securely they would be perpetuated in a context where linguistic divergence was taking place, and it would be reflected in the amount of time taken for competing forms and structures to develop or be grammaticised. If the early stages of Proto Oceanic dispersal were relatively rapid, then perhaps there was not enough time for overt relativisation strategies to develop in higher-order daughter languages before they diversified, but in the later stages there has been enough time for such innovations to occur in closely related groupings of extant languages.

Whether this somewhat informal explanation has merit or not, the distinct relativising forms and strategies found in current languages apppear to represent a high degree of independent innovation, and there seems little prospect of finding historical sources for these forms as relativisers at higher levels of Oceanic.

4 Direction of grammaticisation

The data show that in order to innovate overt relativisation marking, Oceanic languages tend to have drawn on existing resources, and at least 20 out of the 43 languages in the sample appear to have turned towards deictic particles for this purpose. The number could even be greater than this, because other forms now labelled as relativisers might have arisen in the same way, but with the deictic function subsequently lost.

However, this statement raises again the question posed at the start of the paper: how can we know which of the deictic and relativisation functions is primary for those multifunctional forms which serve in both capacities? Two possible reasons can be given here.

Firstly, grammaticisation involves forms with richer lexical meaning losing some of that content and taking on stuctural or grammatical functions. It is true that deictic particles are usually regarded as already grammatical forms, but it can also be argued that they convey meaning that has some semantic relationship to the real world or to the referents in the discourse. This cannot be said of a relativiser, and the only meaning that can be attributed to such particles has to be expressed in purely grammatical terms, associated with the internal syntax of complex clauses. Therefore, within the general understanding of grammaticisation, it is more feasible that semantic bleaching can turn a deictic particle into a relativiser, than the reverse.

Secondly, it can be argued that languages only ‘need’ a single relativiser, while there are good reasons why languages have paradigms of two or more deictic particles. Therefore, as in the data presented here, where languages have more than one form occuring with a relativising function, and these same forms are also found as

demonstratives, it is more likely that the deictic function was primary. It is unlikely that a language would have a number of relativisers which would take on distinct meanings and functions as deictics.

Other factors and cross-linguistic studies point in the same direction. For example, deictics or demonstratives appear as productive starting points in other grammaticalisation processes in Oceanic, being observed to develop into definite articles or tense markers. The 'well-attested grammaticalization path of demonstrative > definite article' (Radetzky 2004:64) is discussed for Oceanic by Ross (1988:100):

> [s]ince there is a strong tendency for Oceanic languages to treat the middle member of the set [of deictic particles], *a/*na, as 'neutral', e.g. to mark a noun modified by a relative clause, it is conceivable that it also became the common article.

As well, the Proto Oceanic

> demonstratives/spatial deictics **e/*ne* 'near speaker', **a/*na* 'near addressee', and **o/*no* 'distant from both speaker and addressee' … may also have served as adverbial particles indicating an event's proximity to or distance from the present (Ross 1988:374),

and eventually taken on the role of tense markers. What is interesting in this case is that Ross (1988:375) considers that while 'the POC deictics **ne*, **na*, and **no* were used temporally', in many Oceanic languages, their development as tense markers is not something that is reconstructible at the earliest stages, but 'their reflexes became part of the verb phrase independently in different Oceanic languages'. In other words, while it is not possible to reconstruct these forms with the function of tense markers at the level of Proto Oceanic, the grammaticalisation path from Proto Oceanic deictic to tense markers in the modern daughter language is clearly apparent, albeit specific to particular languages or lower-order subgroups.

The grammaticalisation of deictic to relativiser is also identified for other languages, as noted by Hopper and Traugott (2003:202): 'uninflected relativisers … may be derived by grammaticalisation from demonstrative pronouns', and their exemplar language for this case is English, with the form *that* being derived from the Old English demonstrative *þe.*

5 Conclusion

As with multifunctional 'that' in English, many Oceanic languages use the same particle both as a deictic (demonstrative or anaphoric reference marker) and also as a relativiser. The deictic meaning is primary, and the grammatical functions of such particles were subsequently expanded to include the relativising function.

While earlier forms of the deictics may be reconstructible at earlier stages of Oceanic language history, their development into relativisers is not something that can be widely detected as a shared innovation for particular subgroups of Oceanic. Instead, it appears to have happened as a process of independent innovation numerous times, and can only be seen now at the level of small very closely related subgroups or individual modern languages.

It appears that languages without relativisers are under structural pressure to develop them, and that existing deictic forms are highly likely candidates for grammaticalisation into this additional role under well-motivated processes of grammaticalisation that can be observed cross-linguistically.

References

Crowley, Terry. 1991. Parallel development and shared innovation: some developments in Central Vanuatu inflectional morphology. *Oceanic Linguistics* 30(2).179–222.

Early, Robert. 1994. A grammar of Lewo, Vanuatu. PhD thesis. Canberra: The Australian National University.

—— 2002. Lamen. In Lynch, Ross and Crowley, eds, 2002, 671–679.

Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Keenan, E. and B. Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.63–99.

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

Nikolaeva, Irena. 2006. Relative clauses. In Keith Brown, ed. *Encyclopaedia of language and linguistics*, 2nd edition, vol. 10, 501–508. Oxford: Elsevier.

Radetzky, Paula Kadose. 2004. Grammaticalization d'un marquer de définitude en saaroa. In Elizabeth Zeitoun, ed. *Fait des langues: revue de linguistique No. 23–24: Les langues austronésiennes*, 213–230. Paris: OPHRYS.

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

—— 2002. Kairiru. In Lynch, Ross and Crowley, eds, 2002, 204–215.

16 *Nasal strengthening in the Fijian languages*

PAUL GERAGHTY

1 Introduction[1]

The Central Pacific subgroup of Eastern Oceanic, comprising the languages of Fiji, Rotuma and Polynesia, was first proposed and named by Grace (1959, 1967), and has since become generally accepted, even though the supporting data is not strong (Geraghty 1983:352–366, 1996:84). The apparent reason for this is that Proto Central Pacific had a short life, its speakers, presumably the bearers of the Lapita culture who were the earliest settlers of Fiji some 3000 years ago, spreading soon all over the Fiji group (or at least the coasts), then rapidly moving on eastwards and diverging. Shared innovations suggest that Proto Polynesian split off from a part of the communalect chain in Tokalau (far eastern) Fiji, comprising eastern Vanualevu, Taveuni, and Lau (Geraghty 1983:381, 2000), whereas Rotuman split off, probably considerably later, from a part of the communalect chain of the north coast of Vanualevu, though a small number of innovations are shared with Western Vitilevu (Geraghty 1996:85, 2000). Subsequently the Fiji communalect chain was maintained, so much so that at least one important phonological innovation — the simplification of the vowel clusters /ae/ and /ao/ — spread through almost all of the group, even though an important change in the syntax of possession had already divided Fiji into two major subgroups (Geraghty and Pawley 1981).

2 Orthography and sources

Unless otherwise indicated, phonetic values in all data and reconstructions in this paper, regardless of source, are as follows: a, e, f, h, i, k, l, m, n, ñ, o, p, r, s, t, u, w, y, z as written; b [mb], c [ð], d [nd], dr [ndr], g [ŋ], j [č], q [ŋg], v [v] or [ß]; x indicates a velar

[1] It is both a pleasure and an honour to dedicate this paper to my valued friend and colleague of many years, Malcolm Ross, whose meticulous work in both the theory and practice of historical reconstruction I have always admired. Malcolm is, like many of us, a historical linguist by calling rather than simply by profession, and I trust and hope that he will continue unearthing Proto Oceanic and Proto Austronesian phonemes, lexemes, and structures for many years to come! My thanks to Ross Clark, John Lynch, Andrew Pawley, Malcolm Ross, and William Wilson, who have helped in various ways with this paper, but who are not in any way responsible for its content.

fricative, ' indicates glottal stop and vowel length is indicated by a macron. Glosses are composed according to the conventions laid out in Geraghty (1983:8–13).

Fijian data are from my field notes, and all Proto Fijian (PFj), Proto Central Pacific (PCP), Proto Southeast Solomonic (PSS) and Proto Southern Oceanic (PSO)[2] reconstructions are my own. For other sources, refer to footnote 3. Irregular reflexes are followed in brackets by a statement of the irregularity, or by a form preceded by a double asterisk which shows the hypothetical regular reflex, or the relevant part of it. Rotuman data are from Inia et al. 1998, but umlaut is not marked, since it is predictable in citation forms. No Rotuman form which is clearly borrowed from a Polynesian language is cited; if a form could be a Polynesian borrowing on purely phonological grounds, it is marked thus: (Pn?).

3 Consonant grade in Oceanic languages

It has long been recognised that many Oceanic languages show two (or more) reflexes of Proto Oceanic oral stops, typically a voiceless stop or fricative (‘oral grade’) and a prenasalised stop (‘nasal grade’), and that it is difficult to determine the conditions for these differing reflexes (see Grace 1990 and references therein, and Blust 1996, who refers to a similar array of unrelated instances of apparently random medial prenasalisation in Western Austronesian as ‘pandemic irregularity’). Only in some cases are there explanations, for example, in the Admiralties where such prenasalisation results from cliticisation of the POc article *na (Ross 1988:337–341), in North and Central Vanuatu where such alternations in initial position of verbs mark tense or aspect (Lynch 1975; Tryon 1986), or in most of Eastern Fiji, where prenasalisation of initial apical stops in common nouns is apparently the result of assimilation to a preceding article *na* or *ni* (Geraghty 1983:74–96). All of these are examples of a particular kind of ‘crossover’: prenasalisation, that is, the change of a stop or fricative to a homorganic prenasalised stop. The purpose of this paper is to present evidence for a different kind of crossover which appears to have occurred in the Fijian languages, one that I believe has not been previously recorded for any Oceanic language: the change from a nasal to a homorganic prenasalised stop. Since there seems to be no name for this kind of change, I will use the term ‘nasal strengthening’.

4 Central Pacific consonant correspondences

Geraghty (1986) lists the consonant correspondences among the three major subgroups of Central Pacific. Since this paper is concerned primarily with the fate of the PCP bilabial and velar prenasalised stops stops and nasals *b, *m, *q and *g, I will present below only the relevant evidence (based on Geraghty 1986:291):[3]

---

[2] Note that I use ‘Southern Oceanic’ to refer to the languages of New Caledonia and the Loyalty Islands, as in Geraghty 1989, whereas others have used the same term to refer to the languages of New Caledonia, the Loyalty Islands, and Vanuatu (Lynch, Ross and Crowley 2002:112).

[3] Abbreviations and default sources are as follows: Are - ‘Are’are (Geerts 1970); Ars - Arosi (Fox 1978); Bug - Bugotu (Ivens 1940); Che - Cheke Holo (White 1998); EF - Eastern Fijian; EFu - East Futuna (Moyse-Faurie 1993); EUv - East Uvea; Fij - any Fijian communalect; Haw - Hawai’ian (Pukui and Elbert 1986); k - kind of; Kwa - Kwaio (Keesing 1975); Lau - (Fox 1974); Mao - New Zealand Maori; Mia – Mangaia; Mqn - Marquesan; Ngg - Nggela (Fox 1955); Ntp - Niuatoputapu; Nuk - Nukuoro; Paa - Paamese (Crowley 1992); PCP - Proto Central Pacific; PFj - Proto Fijian; PMc - Proto Micronesian (Bender et al. 2003a,b);

PCP *b > PFj *b, PPn *p, Rot p

Initial:

*bebe 'butterfly, moth' > PFj *bēbē (**bebe), PPn *pepe, Rot *pepe* (Pn?)

*bā 'wall, fence' > PFj *bā, PPn *pā, Rot *pā* (Pn?)

*buto- 'navel' > PFj *buto-, PPN *pito (**puto),[4] Rot *pufa*

Medial:

*kabe 'string' > PFj *kabe 'string from coconut stem', Rot *`ape*

*tubā 'k land crab, *Cardisoma*' > PFj tubā, PPN *tupa (**tupā), Rot *fupa* (**fupā)

*tubu 'grow' > PFj *tubu, PPn *tupu, Rot *fupu*

PCP *m > PFj *m, PPn *m, Rot m

Initial:

*miji 'suck' > PFj *misi, PPn *miti

*mata- 'eye' > PFj *mata-, PPn *mata, Rot *mafa*

*moze 'sleep' > PFj *moze, PPn *mohe, Rot *mose*

Medial:

*kumete 'bowl' > PFj *kumete, PPn *kumete, Rot *`umefe*

*cama 'outrigger float' > PFj *cama, PPn *hama, Rot *sama*

*ñamu 'mosquito' > PFj *ñamu, PPn *namu, Rot *ramu*

PCP *q > PFj *q, PPn *k, Rot k

Initial:

*qiriqiri 'gravel' > PFj *qereqere (**qiriqiri), PPn *kilikili

*qau 'swim' > PFj *qau, PPn *kau, Rot *kau* 'wade'

*qumuqumu 'k crab' > PFj *qumuqumu, Rot *kumkumu*

Medial:

*leqileqi 'k tree, *Xylocarpus'* > PFj *leqileqi, PPn *lekileki, Rot *lekileki* (Pn?)

*waqa 'canoe' > PFj *waqa, PPn *waka, Rot *vaka* (Pn?)

*(y)aqo 'learn' > PPn *ako, Rot *rako*

PCP *g > PFj *g, PPn *g, Rot g/n

Initial:

*gi(czs)a > Rot *nisa* 'mock, jeer'; PNV *gigica 'smile, grin'

*gara 'scream, cry out' > PFj *gara, PPn *gala 'hoarse, snarl'

*guju- 'mouth' > PFj gusu-, PPn *gutu, Rot *nuju*

PMP - Proto Malayo-Polynesian (Blust n.d.); Pn - Polynesian; PNV - Proto North Central Vanuatu (Clark n.d.); PPn - Proto Polynesian (Biggs and Clark n.d.); PSO - Proto Southern Oceanic; PSS - Proto Southeast Solomons; PSV - Proto Southern Vanuatu (Lynch 2001); Puk - Pukapuka (Beaglehole and Beaglehole 1991); PWMc - Proto Western Micronesian (Bender et al. 2003a,b); PWMP - Proto Western Malayo-Polynesian (Blust n.d.); Rar - Rarotongan; Ren - Rennellese (Elbert 1975); Rot - Rotuman (Inia et al 1998); Rpn - Rapanui (Easter Island); Saa - Sa'a (Ivens 1929); Sam - Samoan (Milner 1966); TF - Tokalau Fijian; Tik - Tikopia (Firth 1985); Tol - Tolo (Crowley 1986); Tua - Tuamotu (Stimson and Marshall 1964); WF - Western Fijian. Unless otherwise specified, the default source for Polynesian languages is Biggs and Clark (n.d.).

[4] It is not clear that PPn *pito is in fact irregular, since PCP *u is often fronted to *i in PPn (e.g. *kuli- 'skin' > *kili), under conditions which have yet to be determined.

Medial:

*tagi ‘cry, weep’ > *PFj tagi, PPn *tagi, Rot *fagi*
*taliga- ‘ear’ > PFj taliga-, PPn taliga, Rot *faliga*
*togo ‘mangrove’ > PFj togo, PPn *togo, Rot *fogo*

5 Irregular bilabial consonant correspondences

In Geraghty (1993:358, 375) I noted that three lexical items show an irregular correspondence between Fijian *b and PPn or PCP *m: Fij *basaga* ‘(road +) fork’, PPn *māsaga ‘(pair of) twins’; Fij bū ‘k fish, *Monotaxis grandoculis*’, PPn *mū; and Fij bubu-c ‘chew and suck’ < PCP *mumu-c.

Then in Geraghty (1994:153, 160) I noted the two following PCP fish names:

*(bm)ū ‘*Monotaxis grandoculis*’
PFj *mū, *bū
PPn *mū

*bōsē ‘*Scarus sp.’*
PFj *bōsē
PPn *mōsē (Ntp pōsē, Puk mōyē ‘k wrasse’)

and added that the Niuatoputapu form may be a loan from Fiji, in which case Puk *mōyē* points to PPn *mōsē, another instance of Fijian initial /b/ corresponding with PPN *m.[5]

Taken together, the above observations provide a small body of evidence for an apparently irregular correspondence between Proto Fijian *b and Proto Polynesian (or Proto Central Pacific) *m. The purpose of this paper is to investigate this correspondence, and other similar correspondences, to determine the directionality of change, and whether any conditioning factors can be discerned.

6 Further evidence

In this section, I will present further cognates, internal and external, of the above comparisons, in order to bolster the reconstructions and suggest a direction of change.

1. PCP *ma(j,s)aga ‘forked’

WF, EF, TF *basaga*
PPn *māsaga ‘(pair of) twins’ (note also Puk *vāyaga* ‘twins’ (**māyaga),
Ren *masaga* ‘branching’)
Rot *majaga* ‘forking, bifurcation’

PSV *na-msag ‘crotch, fork’

[5] I also (1994:160) noted PCP *manoko ‘*Blennidae*’, based on PFj *m(ai)noko and PPn *manoko (EFu *manoko* ‘k fish’, Sam *mano`o*, Nuk *manoko* ‘*Gobiidae*’, Ren *manoko*; contrast EUv *panoko*, Tik *panoko*, Mao *panoko* ‘k fish’, Rar *panako* ‘k small fish like minnow’, Mia *panako*, Rpn *pāroko*, Mqn *pāoko*, Haw *pāo`o*), where the Tikopian, East Uvean and all Eastern Polynesian reflexes show an apparent irregular change from PPn *m to *p (as well as some irregularities in the medial consonant). This is however the only item I am aware of that shows this change.

PSS *masaga ‘branch, fork’ (Tol *masaga-*, Lau *mataga*, Kwa *mataga-* ‘crotch, branching place’, Saa *mataga* ‘forked, branched’; note also Bug *base* ‘forked’, Bug, Ngg *baso* ‘twin’, Ngg *bahaga* ‘crotch, fork in tree’ (**mahaga))

(PMP *saga ‘bifurcation, to branch’)

The bulk of the evidence here supports the PCP reconstruction *ma(j,s)aga ‘forked’, and leads to the conclusion that Fijian languages (including Western, Eastern, and Tokalau) have irregularly changed *m to *b in this form, a change also found in Nggela *bahaga* ‘crotch, fork in tree’, but not in any other Southeast Solomonic language.

2. PCP *mū ‘k fish, *Monotaxis grandoculis*'

EF *bū*

TF *bū, mū*

PPn *mū

PSS *mū (Lau *mū* ‘k white reef fish, good to eat’)

(PMP *mu`ug ‘fish sp.’)

What slight external evidence there is supports the reconstruction of PCP *mū. The putative Lau (Solomons) cognate is not identified as to species, but does fit semantically, since *Monotaxis* is indeed a white reef fish, and good to eat. Note that there is a problem with the putative PMP cognate, since PMP glottal stop should be retained in PCP and PPn, but it is not crucial to the PCP reconstruction. The conclusion is again that Fijian languages have irregularly changed *m to *b, though in this case either the change was not followed in part of Tokalau Fiji, or perhaps the apparent reflex there is a borrowing from a Polynesian language.

3. PCP *mumu-c ‘hold in the mouth and suck’

EF, TF *bubu-c*

(PNV *bubu-si ‘puff, blow’)

PSS *mumu ‘hold in the mouth’ (Kwa *mumu* ‘close mouth, (bivalve) close shell’, *mumu-nia* ‘keep in the mouth’, Lau *mumu* ‘close the lips’, Saa *mumu* ‘close up (mouth+)’, Ars *mumu`i* ‘hold in lips, teeth’)

PMP *mulmul ‘hold in the mouth and suck’

There is strong evidence here for PCP *mumu-c, hence another example of PCP *m > Fiji *b. The PNV form, if cognate, is problematic, but cognacy is doubtful on semantic grounds, and an alternative etymology is that it is a reanalysis of a partially reduplicated form of PEO *pus(iu)-Ø ‘puff, blow’.

The final example is:

4. PCP *mōsē ‘k reef fish, *Scaridae* or *Labridae’*

TF *bōsē* ‘small *Scarus* spp.’

PPn *mōsē (Ntp *pōsē* ‘*Scarus* spp.’, Puk *mōyē* ‘k wrasse’)

This comparison is semantically valid since parrotfish (including *Scarus* spp) and wrasses are very closely related families of reef fish. The Niuatoputapu form, as noted above, looks like a loan from Fijian, to which it is geographically very close. Unfortunately, there is no external evidence to strengthen or otherwise the case for PCP initial *m.

7 More comparisons

Further searching has brought to light more comparisons that appear to show the irregular change PCP *m > Fiji *b:

5. PCP *icama 'mate, fellow'

EF *icaba*

PSS *cama- (Lau *dama*, Saa *dama*-, Are *tama* 'in line, in pairs', Ars *dama*-)

6. PCP *jamu '(eat) carelessly'

WF Nadroga *hamu* '(eat) carelessly'
EF *sabu* 'rubbish', 'carelessly', *sabusabu* '(eat) carelessly'
PPn *samu 'eat scraps, eat one food only'
Rot *jamjamu* '(eat) sparingly'

PNV *zamwa 'chew, fibrous residue' (final vowel probably regular)
PSS *dam(iu) 'chew betel', Lau *damudamu* 'eat noisily'

7. PCP *kumu 'hold in mouth'

WF *kumu-t*
EF *kumu-t, kubu-t*
PPn *kumu 'rinse the mouth' (evidence poor), *puku 'hold something in mouth' (metathesis)
Rot *kumu*

PSO *xxum(u) 'dumb'
PSV *a-qumw-i 'hold in the mouth, put in mouth, suck'
PNV *qumi 'hold in mouth', *komo 'chew with mouth shut, keep food in cheek' (**kumu)
PSS *xumu 'hold liquid in the mouth' (Ngg *xumu*, Ars *qumi*)
PMc *kumwu 'have liquid in the mouth'

8. PCP *lama 'murder, destroy'

WF, EF, TF *laba-t*
PSS *lama (Lau *lama* 'cut down at one stroke', *lama-si* 'take life in revenge', Kwa *lama-si* 'cut branches off felled trees', Saa *lama-si* 'chop up and burn felled trees', Are *rama-si* 'chop up tree to prepare garden', Ars *rama* 'destroy', *rama-ri, rama-si* 'slash, chop off')

Che *lalama* 'spoil, treat roughly, violate, damage'

9. PCP *lumi 'fold'

WF *lubi, lobi* (**lumi)
EF, TF *lobi* (**lumi)
PPn *numi (**lumi)

PNV *lulu-mwi 'roll up' (however PNV *lumi 'fold' is suggested by Paa *lumi* 'fold')
PSS *lumi 'fold' (Saa *lumi* 'turn over edge, hem', Ars *rumi* 'fold')
PMc *lumi 'fold'

Che *lumnu* 'pull up, fold up, fold over'

10. PCP *ma(cz)aki 'sickness, disease'

WF, EF, TF *baca* (loss of final syllable irregular; cf WF, EF, TF *macake* 'k disease, thrush' where the final syllable is retained, but with irregularities in the final vowel and the meaning)
PPn *ma(hs)aki
Rot *masa`i*

PNV *masaki 'sick, fever'
PSS *macaki 'ill, have fever' (cf Bug *vahaxi* 'ill', Ngg *vahaxi* 'in pain, ill, have malaria')
PMc *masaki 'be painful, hurt'

11. PCP *m(a,u)laji 'stale'

EF, TF *bulasi*

PNV *malazi 'mouldy, leftover food'
PSS *maladi 'stale' (Saa *maladi*, Ars *maradi* '(fish) smell stale')

12. PCP *soma '(banana) spathe, inflorescence'

EF,TF *soba*
Rot *soma*
PSO *somV (Nyelayu *tõ* (Ozanne-Rivierre 1998))

13. PCP *timu 'light rain'

WF *yatibu* 'drizzle' (*ya-* is a common prefix in Fijian languages, with various functions)
PPn *timu 'squall, rain'

PSO *timi 'water'
PSS *tim(i,u) 'rain' (Moli *tumu* (Tryon and Hackman 1983:276), Ngg *timi* 'drip', Tol *jimijimi* 'drizzle', Are *imi* 'raindrop')

The above data all support the theory that some instances of PCP *m became *b in all or some Fijian languages; and, if PPn *puku is accepted as cognate, there is also one instance of PCP *m becoming Proto Polynesian *p, in addition to the partial change noted in footnote 5, where *manoko > *panoko* in some western and all eastern Polyesian languages. Where there is disagreement within Fiji (6, 7), it is Western Fijian in both instances, plus parts of Vanualevu in (7), that does not participate in the change. In its geographical extent in Fiji, therefore, this change appears to be similar to, though more extensive than, Eastern Fijian Apical Prenasalisation, which occurred in most of Eastern Fiji, and nasal strengthening may perhaps also have been centred on the Koro Sea prestige area of Eastern Fiji (Geraghty 1983:95).

There seems to be no obvious morphological conditioning factor, since no word class dominates. Phonologically, the fact that all instances but one are before non-front vowels is probably not significant, given the small sample and the fact that /mi/ and /me/ sequences are in any case relatively uncommon in Fijian and Pacific languages generally. More significant, perhaps, is the frequency of occurrences before unstressed vowels; only *mū is an exception, while in the case of *mumu-c it could be argued that the second (pre-tonic) *m is secondarily prenasalised after prenasalisation of the first (see footnote 6 below).

Also noteworthy, though not surprising, is that none of the instances of PCP *m > *b occurs in an accent group which contains a prenasalised consonant. This conforms to the general constraint in Fijian languages on co-occurrence of non-identical prenasalised consonants in the same accent group.[6]

Nevertheless, if *m > *b before a non-stressed vowel in an accent group with no prenasalised consonant were a general rule, there is still no obvious explanation for the fact that, say, *cama 'outrigger' remains *cama* in Fijian languages, rather than **caba*, and *ñamu 'mosquito' is reflected as *yamu* or *namu*, rather than **yabu* or **nabu*.[7] Indeed, PCP *cama 'outrigger' and *icama 'mate, fellow' are an enigmatic quasi-minimal pair, with *m remaining *m in the former, but becoming *b in the latter.[8]

8 Velar nasal strengthening

There are also a few instances of 'nasal strengthening' in velar position in Fijian languages. While PCP *g (velar nasal), as demonstrated above, typically remained a velar nasal in Fijian languages, it became a prenasalised velar stop (orthographic *q*) in the following forms:

14. PCP *gase 'k small plant, epiphyte or fern'

EF *qase* 'k epiphyte, *Dendobrium tokai*, k orchid, *Grammatophyllum*'
PPn *gase 'k plant' (definitions include 'fern')

15. PCP *gase 'weak'

EF, TF *qase* '(animate) old'
PPn *gasegase 'weak, feeble', *gahegahe 'out of breath, exhausted'
PMc (PWMc *gasa 'unable to endure further')

16. PCP *gol(i,o)- '(reed, sugarcane +) top end of shoot'

WF, EF *qolo-*
PNV *goli- '(plant) top shoot'
PSS *goli- '(plant) top shoot' (Lau)

17. PCP *laga 'open slightly, raise a little'

EF, TF *laqa*
PPn *laga 'raise up'
Rot *laga* 'raise one side of, lift or open a little'
PNV *laga-i 'lift flat object from surface'
(PMP *laga` 'gape, open wide')

---

6 Non-identical homorganic consonants are similarly disallowed, or disfavoured, within accent groups. Thus in Standard Fijian, while verbs of motion usually take the transitive suffix *-v-*, e.g *lako-v* 'go to/for', *cici-v* 'run to/for', if there is a bilabial in the root, a non-bilabial transitive suffix is selected, e.g. *kaba-t* 'climb', *cabe-t* 'ascend', *voce-r* 'row to/across'.

7 There are many more examples of PCP *m remaining as *m (i.e., not becoming *b) before an unstressed vowel in Fijian languages, e.g. *cumu 'triggerfish', *macawa 'space between', *malumu 'soft', *mataku 'fear', *matu`a 'old', *mavoa 'wounded, injured', *sama 'k tree, *Commersonia*' **.**

8 There is at least one instance of a parallel change in Vanuatu: PNV *mala 'hawk' is realised as *bala* in most languages of Malakula. My thanks to John Lynch for pointing this out.

18. PCP *māgele ‘k small tree’

EF *māqele* ‘k small tree, *Jatropha curcas*’ (this is an early European introduction to Fiji; it is not known what the earlier referent was)
PPn *māgele ‘k tree, *Trema* sp.’
(Rot *maragi* ‘k tree, *Trema cannabina*’)

19. PCP *taga ‘rest on top of something’

WF *taqa* ‘wear, put on’, EF *taqa* (but EF, TF *i-taga-ga* ‘horned top of mast, on which the yard of the sail formerly rested’)
PPn *taga (Haw *kana* ‘horizontal support in house for carrying poles’, Tua *taga* ‘draw up on the shore, beach’, Rar *taga* ‘on top of’)
(PNV *tabwa ‘lie flat on’)

As with bilabial nasal strengthening, the above forms bear witness to an irregular change of nasal to homorganic prenasalised stop, but offer few clues as to what the conditioning may have been. Neither word class nor stress nor vocalic environment appears to correlate significantly, and the only assertion that can be made is that, as with bilabial position, the change is constrained by the phonotactics of Fijian noted above: it does not occur when the other stop in the accent group is velar or bilabial, or prenasalised. Otherwise put, velar nasal strengthening only occurs when the other consonant in the accent group is a non-prenasalised apical consonant.[9] This constraint is nicely illustrated in example (19), where *taga ‘rest on top of something’ became EF, TF *taqa*, whereas its nominalisation *i-taga-ga ‘horned top of mast, on which the yard of the sail formerly rested’ (Geraghty 2001:67) remained *itagaga*, because of the following velar nasal in the same accent group. Nevertheless, there is no obvious explanation for the fact that, for example, PCP *taga ‘rest on top of something’ became EF *taqa*, whereas *caga ‘span’ remained EF *caga*.

The sparse data here do not permit us to say much about the geographical extent of this irregular change, other than to state that it appears to be exceptionless in Fiji (though only two forms, 16 and 19, are reflected in Western Fijian), and not found at all in any Polynesian language or in Rotuman. Since PCP *gw appears not to have been affected (i.e., there are no examples of PCP *gw > *q), velar nasal strengthening probably took place before the *gw > *g change that occurred in most of Eastern Fiji and in Proto Polynesian (Geraghty 1983:367).

8 Conclusion

I have presented evidence here for two irregular sound changes in the Fijian languages I have termed ‘nasal strengthening’, in bilabial and velar position. There is some evidence that bilabial nasal strengthening may have spread out relatively recently from the ‘Koro Sea’ prestige area of Eastern Fiji, as did Eastern Fijian Apical Prenasalisation, but velar nasal strengthening appears to be more ancient, given its more extensive geographical spread. The conditioning factors are as yet undetermined, though both changes are governed by the usual constraints on co-occurrence of consonants in an accent group.

---

[9] Note that, although *s functioned historically as the ‘prenasalised’ counterpart of *c, it is not phonetically prenasalised, nor does it function phonotactically as a prenasalised consonant (Geraghty 1983:90–91).

It is perhaps noteworthy that I have so far found no evidence for parallel changes in apical or palatal position, i.e. *n > *d or *ñ > *z (prenasalised palatal stop). A similar observation was made by Ross (1988:70–71) in regard to lenition in Western Oceanic languages, where the lenition of *p and *k is widespread, whereas that of *t is 'found only in scattered areas'.

References

Beaglehole, Ernest and Pearl Beaglehole. 1991. Pukapukan dictionary manuscript. Unpublished manuscript. Auckland: Pukapuka Dictionary Project.

Bender, Byron W., Ward H. Goodenough, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel and Judith W. Wang. 2003a. Proto-Micronesian reconstructions. *Oceanic Linguistics* 42(1).1–110.

—— 2003b. Proto-Micronesian reconstructions-2. *Oceanic Linguistics* 42(2).271–358.

Biggs, Bruce and Ross Clark. n.d. *Pollex*. Computer file. Auckland: Auckland University.

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

—— n.d.. *Austronesian comparative dictionary*. Computer file. Honolulu: University of Hawaii.

Clark, Ross. n.d. *North and Central Vanuatu: a comparative study*. Computer file. Auckland: University of Auckland.

Crowley, Susan Smith. 1986. *Tolo dictionary*. Canberra: Pacific Linguistics.

Crowley, Terry. 1992. *A dictionary of Paamese*. Canberra: Pacific Linguistics.

Elbert, Samuel H. 1975. *Dictionary of the language of Rennell and Bellona*. Copenhagen: National Museum of Denmark.

Firth, Raymond. 1985. *Tikopia–English dictionary*. Auckland: Auckland University Press/Oxford: Oxford University Press.

Fox, Charles E. 1955. *A dictionary of the Nggela language*. Auckland: The Unity Press.

—— 1974. *Lau dictionary*. Canberra: Pacific Linguistics.

—— 1978. *Arosi dictionary*. Revised edition, with index by Mary Craft. Canberra: Pacific Linguistics.

Geerts, P. 1970. *'Are'are dictionary*. Canberra: Pacific Linguistics.

Geraghty, Paul A. 1983. *The history of the Fijian languages*. Oceanic Linguistics Special Publication 19. Honolulu: University of Hawaii Press.

—— 1986. The sound system of Proto-Central-Pacific. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 289–312. Canberra: Pacific Linguistics.

—— 1989. Proto Southern Oceanic and its relationships. In Ray Harlow and Robin Hooper, eds *VICAL 1/Oceanic languages. Papers from the Fifth International Conference on Austronesian Linguistics*, 141–156. Auckland: Linguistic Society of New Zealand.

—— 1993. Pulotu, Polynesian homeland. *Journal of the Polynesian Society* 102(4).343–384.

—— 1994. Proto Central Pacific fish names. In Andrew K. Pawley and Malcolm D. Ross, eds *Austronesian terminologies: continuity and change*, 141–169. Canberra: Pacific Linguistics.

—— 1996. Problems with Proto Central Pacific. In John Lynch and Fa'afo Pat, eds *Oceanic Studies: Proceedings of the First International Conference on Oceanic Linguistics*, 83–91. Canberra: Pacific Linguistics.

—— 2000. Possession in the Fijian languages. *Sprachtypologie und Universalienforschung* [Berlin] 53(3–4).243–250.

—— 2001. Amazing -a: The suffix that named the Pacific world. *Rongorongo Studies* 11(2).63–74.

Geraghty, Paul and Andrew Pawley. 1981. The relative chronology of some innovations in the Fijian languages. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 159–178. Auckland: Linguistic Society of New Zealand.

Grace, George W. 1959. *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family*. Bernice P. Bishop Museum Special Publication 46. Honolulu: Bishop Museum Press.

—— 1967. The effect of heterogeneity in the lexicostatistical test list: the case of Rotuman. In Highland, Genevieve A., Roland W. Force, Alan Howard, Marion Kelly, Yosihiko H. Sinoto, eds *Polynesian culture history: essays in honor of Kenneth P. Emory*, 289–302. Honolulu: Bernice P. Bishop Museum Press.

—— 1990 'Consonant grade' in Oceanic languages. In J.H.C.S. Davidson, ed. *Pacific Island languages. Essays in honour of G.B. Milner*, 41–49. London: School of Oreiental and African Studies/Honolulu: University of Hawaii Press.

Inia, Elizabeth K, Sofie Arntsen, Hans Schmidt, Jan Ressel and Alan Howard. 1998. *A new Rotuman dictionary*. Suva: Institute of Pacific Studies, University of the South Pacific.

Ivens, Walter G. 1929. *A dictionary of the language of Sa'a (Mala) and Ulawa/South-East Solomon Islands*. Melbourne: Oxford University Press/Melbourne University Press.

—— 1940. *A dictionary of the language of Bugotu*. London: Royal Asiatic Society.

Keesing, R.M. 1975. *Kwaio dictionary*. Canberra: Pacific Linguistics.

Lynch, John D. 1975. Oral/nasal alternation and the realis/irrealis distinction in Oceanic languages. *Oceanic Linguistics* 14(2).87–99.

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

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

Milner, G.B. 1966. *Samoan dictionary*. Oxford: Oxford University Press.

Moyse-Faurie, Claire. 1993. *Dictionnaire futunien-français*. Paris: Peeters.

Ozanne-Rivierre, Françoise. 1998. *Le Nyelâyu de Balade (Nouvelle-Calédonie)*. Langues et Cultures du Pacifique 12. Paris: Peeters.

Pukui, Mary Kawena and Samuel H. Elbert. 1986. *Hawaiian dictionary: Hawaiian–English, English–Hawaiian*. Honolulu: University of Hawai'i Press.

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

Stimson, Frank and Donald Marshall. 1964. *A dictionary of some Tuamotuan dialects of the Polynesian language*. The Hague: Martinus Nijhoff.

Tryon, Darrell T. 1986. Stem-initial consonant alternation in the languages of Epi, Vanuatu: a case of assimilation? In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 239–258. Canberra: Pacific Linguistics.

Tryon, Darrell T. and Brian D. Hackman. 1983. *Solomon Islands languages: an internal classification*. Canberra: Pacific Linguistics.

White, Geoffrey M. 1988. *Cheke Holo (Maringe/Hograno) dictionary*. Canberra: Pacific Linguistics.

17 *On reconstructing pronominal proto-paradigms: methodological considerations from the Pama-Nyungan language family of Australia*

HAROLD KOCH

1 Introduction[1]

1.1 Background

This study emerges from a project involving the comparison of Australian languages suspected of belonging to the Pama-Nyungan (PN) family,[2] with a view to (a) deciding whether there is enough comparative evidence to justify such a genealogical grouping, (b) reconstructing the proto-grammar of the ancestral language, Proto Pama-Nyungan (pPN), and (c) compiling an etymological database of cognate forms that support the reconstructions. One of the main aspects of the putative proto-grammar involves the case inflection of nominals; i.e. nouns, personal pronouns, deictic-anaphorics, and interrogative-indefinites.

---

[1] An earlier version of this presented at the annual conference of the Australian Linguistic Society, Newcastle NSW, 28 September 2003, under the title 'Particles and proto-paradigms: extending the catchment area for the reconstruction of pronominal inflection'. I am happy to present this version to Malcolm Ross, who has pioneered the study of pronouns as evidence for genetic relations among the Papuan languages (Ross 2001, 2005). It should be noted that, whereas Ross uses the term 'pronoun paradigm' in the sense of a set of forms expressing different person-number values, I am concerned here with case paradigms of individual pronouns that express a single person-number combination or a demonstrative or anaphoric (or even interrogative) function. I acknowledge helpful feedback from the editor and an anonymous referee.

[2] The language family, as proposed in the early 1960s by Hale, O'Grady, and Wurm (see Wurm 1972) and revised by Blake in the late 1980s (see Blake 1988), supposedly encompasses most of the languages of the Australian mainland, except for most of those of north-western and north-central Australia. I largely leave out of consideration the poorly-documented PN languages of the southeast in this preliminary attempt at a detailed comparison — while admitting that they should eventually be explored for their contribution to the reconstruction of these paradigms.

1.2 Issues

Several issues confront us in the reconstruction of such inflection. First, there is a substantive issue: what was the **system** of case inflection of the protolanguage? In other words, what are the reconstructable case values and what are their exponents (which may be conditioned by phonological, morphosyntactic or semantic features of the lexical stems)? Second, there is a methodological issue: what **methods** should be used to reconstruct the system? Finally, there is a practical issue (which also involves methodology): where do we look for the **evidence** on which to base our reconstructions?

1.3 Aims and outline

In §2 I survey the history of reconstructing PN pronominal paradigms, especially those for the monosyllabic vowel-final stems *nhu-, *pa-, and *ngu-. In the process I make explicit the methodological arguments and assumptions that are employed. I characterise the procedures that are used for the comparison of: inflectional patterns, inflected words only, inflected words plus cognate stems, and inflected words plus cognate particles. In §3 I add further data relevant to *pa- and *ngu-, with particular attention to the evidence available from particles that can plausibly be interpreted as erstwhile inflected forms. In §4 I summarise the substantive results and comment on their implications for language change and reconstruction.

1.4 Note on transcription

In this paper all forms are presented in a uniform orthography, which uses the conventions set out in Table 1. Note that some languages lack the dental vs. prepalatal contrast: in such languages the laminal consonants are represented with the *Cy* symbols. Furthermore, some languages lack the apical contrast; in such cases the sole apicals are represented with symbols otherwise used for alveolars. Note further that many languages lack the vowel length contrast. A few languages have a voicing contrast; for these languages the symbols *b*, *d*, *g*, etc. are used.

Table 1: Typical Pama-Nyungan (Australian) phoneme inventory

	Peripheral		Laminal		Apical	
	Labial	Velar	Dental	Prepalatal	Alveolar	Postalveolar
Obstruent	*p*	*k*	*th*	*ty*	*t*	*rt*
Nasal	*m*	*ng*	*nh*	*ny*	*n*	*rn*
Lateral			*lh*	*ly*	*l*	*rl*
Tap/Trill					*rr*	
Approximant	*w*			*y*		*r*
Vowel long	*uu*		*ii*		*aa*	
Vowel short	*u*		*i*		*a*	

1.5 Summary of previous reconstructions of Pama-Nyungan case paradigms

The pPN inflection of disyllabic vowel-final common nouns and of some pronominals is discussed in Koch (2003b). There it is argued that there is comparative support for the reconstruction of a system shown in Table 2, which presents reconstructed suffixes for the cases of certain classes of stems as well as specific pronominal stems.[3] (Here N stands for noun and Pn for pronoun.)

Table 2: Summary of PN protoparadigms (from Koch 2003b)

Stem	ERG	NOM	ACC	DAT	LOC	ABL
N *CV(C)CV-*	*-ngku*	*Ø*	*Ø*	*-ku*	*-ngka*	*-ngu*
Pn *CVn-*:	*-tu*	*Ø*	*-a*	*-u*	*-ta*	
2Sg *nyun	*nyuntu	*nyun	*nyuna	*nyunu		
3SgF *nhan	*nhantu	*nhan	*nhana	*nhanu		
'who/what'	*ngaantu	*ngaan	*ngaana	*ngaanu		
Pn *CV-*:	*-lu*	*Ø*	*-nha*	*-ngu*		
3SgM *nhu	*nhulu	*nhu	*nhunha	*nhungu		
Pn 1Sg *nga-	*ngathu	*ngay	*nganha	*ngatyu		

Note that personal pronouns tend to lack a Locative case form. Further, in many languages the Dative suffix for personal pronouns is different from that of other pronominals. The Accusative inflections likewise often differ, with vowel-final stems stems referring to persons taking *-nha and non-personal stems taking *-Ø. Koch (2003b) also gave tentative reconstructions for the deictic stems *ngu and *pa — interpreted as instances of the 'Pn CV-' pattern of Table 2. The evidence for these inflections and their implications for a proto-paradigm will be discussed in greater detail in §3 below. It must be asked whether the comparative data has sufficient geographical breadth to justify reconstruction of an inflectional paradigm. It will be shown that only a partial paradigm can be securely reconstructed for pPN, even after we have expanded the catchment area for available cognate forms.

2 Methods used previously for reconstructing pronominal inflectional paradigms

2.1 Method of projecting patterns

The first method we shall consider for the reconstruction of case inflections can be characterised by the description in (1).

(1) Compare rules/patterns of the relevant languages and project these back to the protolanguage.

---

[3] Abbreviations for cases are as follows: ERG - Ergative (transitive subject); NOM - Nominative (intransitive subject); ACC - Accusative (direct object); DAT - Dative (which marks a range of functions including recipient, beneficiary, purpose, and sometimes possessor); LOC - Locative; ABL - Ablative (source); OBL - Oblique (often stem/base used for cases other than ERG, NOM, ACC, especially with nominals referring to humans); ABS - Absolutive (a cover term for identical NOM and ACC).

This approach has been widely used with reference to the reconstruction of Ergative and Locative allomorphs (Blake 1979b, 1988), Dixon (1980, 2002), Sands (1996). A widespread PN pattern of case allomorphy is shown in Table 3.

Table 3: Widespread PN pattern of case allomorphy

	/ C_	/ V_
Ergative	*-Tu*	*-ngku ~ -lu*
Locative	*-Ta*	*-ngka ~ -la*

(T = stop homorganic with C)

The choice of *-ngku/a* versus *-lu/a* is determined, in different languages, by one or more of: (a) the number of syllables, (b) common versus proper status of the noun, (c) the grammatical class (pronominal versus noun). Sands emphasises the third factor in her reconstruction of the distribution of the pPN Ergative allomorphs, positing *-lu as the allomorph appropriate to pronominals.

> in many languages any of the pronouns, demonstratives or interrogatives with final vowel take allomorph *-lu* [for the Ergative]. (Sands 1996:18)

This reconstruction is supported by cognate paradigms only for the 3Sg personal pronoun, for which Sands reconstructs ERG *NHulu beside ACC *NHuNHa.[4]

> For the demonstratives there are no reconstructable Pama-Nyungan forms with final *-lu*, but *-lu* appears in different languages on quite different demonstrative forms. It therefore appears that there is a tendency for demonstrative forms to take the allomorph *-lu*. Interrogatives also have a tendency to take the allomorph *-lu* [of the Ergative]. (Sands 1996:22)

By the method described in (1), comparing patterns and projecting them backwards in time, Sands therefore reconstructs a pPN pattern as shown in Table 4:

Table 4: pPN pattern of Ergative allomorphy with post-vocalic stems (Sands 1996:22)

	common nouns	pronominals (plus non-common nouns)
Ergative	*-ngku*	*-lu*

The disadvantage of this kind of pattern-based argumentation is that it overlooks the possibility that the languages may have independently extended a pattern of case-marking to particular pronominal stems that had a different inflection at an earlier stage.[5]

---

[4] Here *NH* stands for a laminal nasal that is reflected as a dental *nh* in languages having a laminal contrast but by an unmarked laminal *ny* in languages with a single laminal series, according to the conventions established in Dixon (1970).

[5] Compare Meillet's (1967:65–66) example of parallel innovation in different Indo-European languages that independently extended the use of *-m* to mark 1Sg subject on verbs that previously had a different inflection.

2.2 Method of comparing whole inflected words

A more traditional method of reconstructing inflection is described in (2).

(2) Compare whole inflected words belonging to paradigmatic sets in the related languages, reconstruct protoforms by the Comparative Method, then perform morphological analysis on the members of the proto-paradigms.

This is a method of morphological reconstruction described by Anttila (1972:351) as 'applied phonology' (cf. Evans 2003).

The approach described in (2) is essentially what is used in Dixon's (1980, 2002) discussion of the inflection of CV- pronominal stems — supplemented by a further method that will be discussed in §2.3 below. Tables 5, 6, and 7 present some of Dixon's evidence for his 'suggested reconstructions'.[6] From Table 5[7] it can be seen that a monosyllabic root *nhu-/nyu-*, indicating the 3Sg (Masculine if a gender distinction is made) pronoun, is analysable on the basis of at least two contrasting inflectional forms in the same language — from a considerable number of languages of Queensland (the Wik languages of western Cape York Peninsula, Warrgamay from the Rainforest area, Warrungu and Gugu Badhun of the Mari subgroup, Wangkumarra and Pitta-Pitta of the Karnic subgroup,[8] plus generally languages of the Warluwarric subgroup). The expected suffix-less Nominative form *nhu/nyu* occurs with an increment in several languages: *-ya* in Wangkumarra, *wa* (≺ *-pa?) in Pitta-Pitta, Warluwarric languages, and further in two closely related languages of the New South Wales central coast.[9]

Table 5: 3Sg(M) *nhu- (source Dixon 1980, 2002)

Language	Gloss	ERG *nhulu	NOM *nhu	ACC *nhunha	(Oblique)	Reference
Proto Wik	3Sg	*nhulu* [a]		*nhunha*	*nhungu-*	2002:316
Warrgamay	3Sg	*nyula[ngka]*			*nyunga* [b]	1980:358
Warrungu	3Sg	*nyula* [a]		*nyunya*	*nyungu* [c]	1980:359
Gugu Badhun	3Sg	*nhula* [a]		*nhunha*	*nhungu* [c]	2002:306
Wangkumarra	3SgM	*nhulu*	*nhiya*	*nhunha*	*nhu(ng)-*	2002:305, 313
Pitta-Pitta	3SgM	*nhulu-*	*nhuwa-*	*yinha-*	*nhu-*	2002:305
pWarluwarric	3SgM		*yu(wa)			2002:306, 461
Kattang	3SgM		*nyu(wa)*			2002:461
Awabakal	3SgM		*nyuwa*			2002:461

[a] These forms function as both Ergative and Nominative cases.
[b] This form functions as Nominative and Accusative cases.
[c] These forms function as Genitive cases.

---

[6] In Dixon (1980) these reconstructions are ascribed to a 'proto-Australian'. In Dixon (2002) they are only 'earlier forms', not ascribed to any particular node of a genealogical construct, since he doubts the possibility of demonstrating any genetic grouping (including pPN) at a level higher than some 30 fairly obvious subgroups and rejects the idea that an articulated family tree can be constructed for the Australian languages.

[7] Material in square brackets is extraneous to the comparison.

[8] Karnic is not accepted as a genetic grouping by Dixon (2002). For an alternative view see Bowern (2001).

[9] Phonological changes assumed by the comparisons in Table 5 include: *u* > *a* finally in the Warrgamay, Warrungu, Gugu Badhun Ergative; *uya* > *iya* in the Wangkumarra Nominative, *ny* > *y* initially in pWarluwarric (and in the Pitta-Pitta ACC).

Table 6 gives Dixon's evidence for the inflection of a monosyllabic root *ngu-*, which is a 3Sg pronoun in the Flinders Island language of Cape York Peninsula but a remote demonstrative ('yon') in languages from the far southwest of Australia. In these languages the contrast between two or more inflected forms justifies the positing of a monosyllabic stem plus inflections. The languages agree on an Ergative *ngulu (with Tharrgari *l > *d*). The Oblique forms do not agree. Dixon's Accusative in *-nha* is reflected only in the two western languages, where it is actually the Nominative form. Locative forms are not given.

Table 6: 'distant' Demonstrative *ngu- (source Dixon 1980:360, 2002:306)

Language	Gloss	ERG *ngulu	NOM *ngu	ACC *ngunha	(Oblique)
Flinders Is.	3Sg	*ngulu* [a]			*ngungu-*
Thalanji	Dem remote	*ngulu*		*ngunha* [b]	
Tharrkari	Dem remote	*ngudu*		*ngunha* [b]	*ngurnu* [c]

[a] This functions as both Ergative and Nominative cases.
[b] These function as Nominative case.
[c] This functions as a Genitive/Dative case.

Note that we have the same pattern of inflection for *ngu as we saw in Table 5 for *nhu. Dixon (1980:361) treats the *ngu-* forms as cognate with *nhu-, and derives them by a sound change *nhu* > *ngu*. This allows him to reconstruct a protoparadigm of *nhu with Erg *-lu, Nom *-Ø, Acc *-nha (Dixon 1980:362), a judgement repeated in Dixon (2002:305):

> [There is] enough [data] to suggest an earlier paradigm (for some of the languages in groups A-Y, WA-WM):[10]
>
> | (65) | S | A | O |
> |---|---|---|---|
> | 3sg.(m) | *nhu* | *nhu-lu* | *nhu-nha* |
> | 3sg.f | *nhan* | *nhan-du* | *nhan-nha* |

An etymological relation between *nhu and *ngu can not, however, be maintained. In the first place, the putative sound change would have to be supported with further cognates. A more serious problem is the fact that both roots co-occur (sometimes in extended form) in a number of languages. Warrungu in the northeast has *nyula* '3Sg' versus *nguna* 'that' and *nguni* 'there' (Tsunoda 2003). Martuthunira and several adjacent languages of the Pilbara region of Western Australia have paradigms of both *ngu and *nhu as distal and proximal demonstratives respectively reflecting inflected forms ERG *ngulu and *nhulu, NOM (not ACC) *ngunha, Dative *ngurnu and *nhurnu, and LOC *ngula and *nhula, according to Dench (2007:226–227).[11] Methodologically we must keep the *nhu- and *ngu- paradigms

[10] This is the way the Pama-Nyungan languages are designated in Dixon (2002). Dixon's S, A, and O correspond to the Nominative, Ergative, and Accusative cases respectively.

[11] Dixon (1980:359ff.) implies that the *nhu- paradigm is found only in languages in the eastern part of Australia (including many not given here, see Dixon 2002: 461) and that the *ngu- paradigm is restricted to western languages — plus a single language (Flinders Island) of Cape York Peninsula. A 3Sg stem based on *ngu-*, however, occurs in the Central NSW languages — *nguru* in Gamilaraay and *nguu* in Yuwaalaraay/Yuwaalayaay; cf. Ash et al. (2003), Austin et al. (1980), Williams (1980). A demonstrative stem *ngunu*, *nguna-* 'this, here' further occurs in the closely related Yuin languages Gundungurra (Mathews and Everitt 1900) and Ngunawal (Mathews 1904) in south-eastern New South Wales. See §3.3 for examples of derivatives (if not paradigms) of *ngu-* in eastern and Yolngu languages.

separate, while noting the similarity of their inflectional patterns. Dixon's comparative evidence, then, does not provide a sufficient geographical spread of languages to reconstruct either paradigm to pPN; but when Dench's Pilbara (western) data is combined with the (eastern and east-central) data in Table 5, the evidence for *nhu-lu is probably sufficiently widespread to justify reconstruction to pPN of Dixon's Ergative, Nominative, and Accusative forms. There are many further eastern languages which support this paradigm and provide evidence for a high-level (if not pPN) reconstruction of a Dative/Oblique *nhungu. Further data bearing on the reconstruction of a paradigm for *ngu- will be explored in §3.2.

Before proceeding further, it is worth reflecting on the methods used to reconstruct the paradigms of *nhu and *ngu. We first apply to inflectional forms within particular languages the method of synchronic morphological analysis, described in (3). This establishes the forms as paradigmatic sets in each language.

(3) A contrast *Ax, Ay, Az* in a single language justifies the identification of a stem *A* and inflections *x, y, z*; i.e. *A-x, A-y, A-z*, where the hyphen indicates a morpheme boundary.

We then apply the comparative method, (2), to these inflected forms to reconstruct a set of proto-words *Ax, *Ay, *Az analysable in the synchronic morphology of the proto-language as constituting an inflectional paradigm; i.e. a stem *A followed by inflectional suffixes *-x, *-y, *-z.

This method has the advantage over the pattern-comparison method in (1) of comparing whole words that are etymologically related. A methodological limitation is that regularly inflected forms, even if shared across languages, may be new within their language, having been created analogically according to synchronically productive inflectional processes. A safeguard is to put more reliance on comparable inflected forms which are not regular within their synchronic language systems. A second, practical, limitation is that too few comparable inflected forms may be available to justify a solid reconstruction. Where the weight of evidence for reconstructed forms is too light, confidence in the reconstructions is reduced, even though they may be historically correct.

2.3 Method of comparing stems with inflected words

A third monosyllabic pronominal stem discussed by Dixon is the 'mid-distant' Demonstrative ('that') *pa, which in some languages is reflected as a third person pronoun. Dixon's supporting data is presented in Table 7 (where SG means a sub-group), arranged according to his 'suggested proto-forms'. There is not much direct evidence for an inflecting monosyllabic **root**. The best direct evidence would appear to be the contrasting forms of Karlamay, from which we can infer *pa* plus *-lu* ERG, *-la* LOC. Rather, the comparative inflectional forms are supplemented by another kind of data — **alternate inflecting stems**. Thus in Payungu beside the regular inflecting stem *pala-* there occurs an irregular NOM *panha.* Likewise in Ngarluma the regular stem is *palu-*, but the ACC *parnumpangu* appears to be based on a stem *parnu-,* which occurs elsewhere in western languages as a Dative. Following Dench's (in preparation) study of pronominals in the languages of the Pilbara, Dixon concludes that paradigms have been restructured 'with an original case-inflected disyllabic form being taken as the new root', and suggests proto-forms as given in the second line of Table 7 (Dixon 2002:335).

Table 7: 'mid-distant' Demonstrative *pa- (source Dixon 2002:335-336)

Language	Area[12]	Gloss	ERG *pa-lu	NOM *pa	ACC *pa-nha	LOC *pa-la	Other
Dyirbal	NE	that				*pala-*	
Dharumbal	E	that			*panha*		
Kalkatungu	C	that		*paa-*			
Wirangu	S	that			*panha*	*pala*	
Mirning	SW	3Sg			*panha*		
Karlamay	SW	that	*palu*			*pala*	
Nyungar	SW	3Sg				*pal-*	
Kartu SG	W	3Sg	*palu-*			*pala*	
Payungu	W	that			*panha* [a]	*pala-*	
Thalanji	W	that				*pala-*	
Ngarla	W	3Sg	*palu-*				
Panyjima	W	that			*panha-*		
Ngarluma	W	that	*palu-*				*parnumpangu* [b]
Nyamal	W	3Sg	*palu-*				
Marrngu SG	NW	that				*pala*	

[a] This functions as a Nominative case.
[b] This functions as an Accusative case.

This implicitly involves supplementary methodological assumptions and procedures as follows. The method of Internal Reconstruction, described in (4), is first applied in individual languages.

(4) Alternative inflecting stems displaying formal and semantic similarities (*Ax-* and *Ay-*) in the same language justify reconstructing for an earlier stage of the language a paradigm in which the common element is a stem and the differing elements are inflections; i.e.: *A-x, *A-y.

This internally reconstructed paradigm is then compared to paradigms that actually occur in other languages, and the earlier hypothesised inflections are compared to the actual inflections of the other languages. This is the supplementary method described in (5).

(5) From cross-language comparison of a cognate inflected form (*A-x*) with an inflectional stem (*Ax-*) reconstruct an inflected form (*A-x).

A good example of the operation of these procedures is the comparison of Wirangu *panha(-)* and *pala*(-) of Table 7, which are alternative deictic/3Sg inflecting stems rather than members of the same paradigm (Hercus 1999:83), with cognate inflected forms in Kaurna and Parnkalla (see Table 8).

[12] In place of Dixon's letter-number codes, I indicate the broad geographical region where the language was spoken. Abbreviations used here and in Tables 16 and 20 are: C centre, E east, N north, NE northeast, NW northwest, S south, SE southeast, SW southwest.

Geoff O'Grady (1990:210f.), as part of his Pama-Nyungan etymologies, reconstructed a pPN deictic stem *pa 'that; that way, thither'. His cognates include 'suffixed forms' whose suffixes are glossed with case meanings: Ergative, Accusative, Locative, Oblique (plus a rare Elative). Some of his supporting data is given in Table 8, arranged according to his reconstructed suffixed forms.[13] The first few lines show languages for which O'Grady provides reflexes of more than one proto-form; i.e. members of a potential paradigm. The South Australian languages Kaurna and Parnkalla, of the Thura-Yura subgroup, show inflected forms of a monosyllabic root *pa-*. O'Grady also noted the etymological significance of alternative stems in Ngarluma and Payungu. We can also see the cognacy of Ngarluma *parnu-* with the Oblique inflection/stem of the Thura-Yura languages. Table 8 also gives further instances of *pala* and *panha.*

Table 8: pPN Deictic 'that; that way, thither' (source O'Grady 1990:210f)

Language	Absolutive *pa	Ergative *pa-lu	Locative *pa-la	Accusative *pa-nya	Oblique *pa-rnu
Kaurna	*pa* [a]	*padlu*	*padla[tya]*		*parnu* [k]
Parnkalla		*padlu*		*panha*	*parnu[ntyu-]*
Ngarluma		*palu* [c]			*parnu[mpangu]* [l]
Payungu			*pala* [d]	*panha[..]*	
Mirniny				*panha[rtu]*	
Nhanta				*anha*	
Yingkarta				*panya*	
Panyjima				*panha* [j]	
Nyangumarta			*pala[rri]* [e]		
Dyirbal			*pala-* [f]		
Nyungar		*pal*	*palay* [g]		
Ritharngu			*bala* [h]		
Kala Lagaw Ya	*pa* [b]		*balanya* [i]		

[a] 3Sg
[b] (motion away)
[c] Nominative case
[d] 3Sg
[e] thither
[f] there (Noun Marker)
[g] look out!
[h] that way
[i] like this
[j] 3Sg
[k] 3Sg Possessive
[l] Accusative case

Dench (2007, 2008, In preparation), in his comparative study of pronouns and demonstratives in the languages of the Pilbara, makes extensive use of the method of comparing alternative stems with one another and comparing these stems with inflected words. His results include the reconstructed paradigms given in Table 9.

[13] Actually, the forms I have presented under *pa-la LOC are listed by O'Grady either under *pa+la+rri 'that+LOC+ALL' or, puzzlingly, under *pa+lu 'that+ERG'.

Table 9: Dench's reconstructed Pilbara demonstrative forms (source Dench 2008)

	Proximal *nhu-*	Mid-distal *pa-*	Distal *ngu-*
NOM/ACC	*nhay(i)	*panha	*ngunha
ERG	*nhulu	*palu	*ngulu
LOC	*nhula	*pala	*ngula
OBL	*nhurnu	*parnu	*ngurnu

The method of comparing inflectional stems with inflected forms increases the amount of available comparative data and thus improves the prospects of reconstructing an inflectional paradigm. One additional procedure increases the catchment area for comparison even further.

2.4 Method of comparing uninflected forms (particles) to inflected forms

I would like to call attention to the cognates presented by O'Grady which I have displayed in the last three lines of Table 8 and discuss their implications for the methodology of reconstruction. These reflexes are neither case forms nor even pronominal stems (whether deictic or personal pronouns), but rather non-inflecting particles.[14] The assumption regarding language change is that certain members of an inflectional paradigm may be reanalysed as independent lexemes no longer related to the rest of the paradigm; this process can be called deparadigmatisation (cf. Trask 2000:86). The supplementary method that follows from this process is described in (6). This consists of comparing unanalysable, non-inflecting words to corresponding forms from other languages that are analysable in terms of a root plus an inflection, as well as to inflecting forms that according to (5) may continue erstwhile inflections. These non-inflecting words are hence interpreted as erstwhile inflected forms and are used to provide support for the reconstruction of proto-paradigms.

(6) Compare non-inflecting words (= particles), *Ax*, with corresponding inflected forms (*A-x*) and inflecting stems (*Ax-*).

This method has the virtue of using forms which are not synchronically analysable and therefore not liable to have been created by the prevalent morphological patterns of the current language. A possible weakness of this method, however, is that justifying an etymological connection between particles and their erstwhile paradigms is more difficult than comparing obvious cognate inflected lexemes.

O'Grady (1990:213) also includes in his cognate comparisons what he calls cliticised forms of the deictic *pa, citing the Nyangumarta Relative Clause Marker *=pa* and the Ngiyampaa Subordination Marker[15] *-pa*. These forms, which were not given in Table 8, are shown in Table 10. These *pa* cognates, in addition to being neither inflected forms nor inflecting lexical stems, are phonologically bound forms (i.e. clitics) and have a specialised grammatical function.

---

[14] Likewise Dench (2007:223) observes, as a footnote to his discussion of demonstrative paradigms in languages of Western Australia: '[i]n some languages, members of the original paradigm may surface as non-inflecting particles with a range of functions.'

[15] For the details of the syntax of the Subordination Marker *=pa*, see Donaldson (1980).

Table 10: Further comparative data for *pa- (source O'Grady 1990:213)

pPN	*pa	'that' (ABS = NOM/ACC)
Nyangumarta	*=pa*	Relative Clause Marker
Ngiyampaa	*=pa*	Subordination Marker

The methodological assumption being made here is that the uninflected (or zero-suffixed) Absolutive member of a paradigm has become an invariant particle with its own particular function. The comparative procedure employed here is a special case of (6), and can be described in terms of (7).

(7) Compare a non-inflecting word, i.e. particle, (*A*) to a corresponding zero-inflected member of a paradigm (*A-x* where *x* = Ø). (Here zero, the contrastive absence of affix, is the exponent of an inflectional property.)

This (final) procedure allows for particles to be included in the comparanda for reconstructing proto-paradigms. This procedure, however, requires etymological justification. Given the few phonological segments involved in most clitics, there is the risk of catching short forms which are not really cognate. (For example, there are many instances of particles *pa* in Australian languages, and it is not easy to discern which ones may be related to a former inflecting stem.) The expanded catchment area for cognates that is opened up by the inclusion of particles will be exploited in §3 to add to the data relevant to the reconstruction of the pronominals *pa and *ngu.

3 Further data for the reconstruction of *pa- and *ngu

3.1 Limitations of previous results

The pronominal reconstructions that have been mentioned in the preceding discussion are summarised in Table 11. It needs to be asked to what extent these can be reconstructed to pPN. In the absence of an agreed subgrouping at the highest level within PN, we require a geographical distribution of cognates that is broad enough to preclude the forms occurring only in a set of languages that may constitute a subgroup (cf. the approach in Koch 2003b).

Table 11: Reconstructed forms mentioned in foregoing discussion

Gloss	Stem	ERG	NOM	ACC	LOC	DAT/OBL
3SgM	*nhu-	*nhulu	*nhu	*nhunha	*nhula	*nhungu/*nhurnu
'distant'	*ngu-	*ngulu	*ngu	*ngunha	*ngula	*ngurnu
'mid-distant'	*pa-	*palu	*pa	*panha	*pala	*parnu

The present results involve the following limitations in distribution. There are few attested instances of forms which would justify the reconstruction of a (bare) Nominative case form for any of the three stems. The OBL form *nhungu is found only in eastern languages, while the OBL forms in *-rnu are found only in western languages. Inflected forms of *ngu are found only in the west, except for the Flinders Island language, whose

forms could conceivably be reflexes of *nhu-, since the languages of Cape York Peninsula generally have the *nhu- stem for 3Sg. Presumed inflectional reflexes of the stem *pa- are found in Eastern languages only in Dyirbal *pala*, Dharumbal *panha*, and possibly Kala Lagaw Ya *balanya*. These limited results highlight the desirability of finding supplementary data, which in turn requires the application of supplementary methods.

It is desirable to try to reconstruct whole paradigms for each of the pronominal stems separately. This is because we cannot expect a priori that all stems will have exactly the same inflectional pattern. Pronominals (and other nominals) that characteristically refer to humans in many Australian languages have different Dative/Oblique and Accusative inflections from those that have non-human referents (including 'who' versus 'what'). Thus we cannot rely on a general inflectional pattern and assume, for example, that because 3SgM *nhu has Accusative *-nha* so will deictics *ngu and *pa. Or that deictics will share an Oblique *-ngu* with the 3Sg personal pronoun *nhu-ngu. Further we can expect that the inflectional pattern of a stem will probably be altered if the stem changes its function from that of a deictic to a 3Sg personal pronoun — which has been a development with both *ngu- and *pa-.[16] In the following sections I will broaden the comparative materials for both of these stems, organising the evidence according to whether it comes from: (a) inflected forms of a monosyllabic root, (b) disyllabic stems which are themselves reflexes of former inflected forms, or (c) particles which are reflexes of former inflected forms.

3.2 Further data on *pa-

3.2.1 Inflections of monosyllabic root

The best direct evidence for inflection of a monosyllabic root *pa- comes from the Thura-Yura (T-Y) languages of South Australia. Table 12 gives forms cited by Simpson and Hercus (2004). Note that the Kaurna and Parnkalla forms are given in the phonetic spelling of their 19th century sources. Several languages show the effects of prestopping of a lateral and/or nasal consonant. In Adnyamathanha initial *p* has lenited to a fricative represented by *v*. Note further that the Dative forms are given as Genitive in my source.

Table 12: Inflected monosyllabic root *pa-

	Gloss	ERG	NOM	ACC	DAT
Kaurna	3Sg	*padlo*	*pa*	*pa*	*parnu*
Parnkalla	3Sg	*padlo*	*panna, =wo*	*panna*	*parnüntyuru*
Narangga	3Sg	*palu*	*pa*	*pa*	*panu*
Kuyani	3Sg	*palu*	*panha*	*panha*	*pardnityarunha*
Adnyamathanha	3Sg	*valu*	*vanha, =wa*	*vanha, =wa*	*vardnuntyaru*
pT-Y	3Sg	*palu	*pa(nha)	*pa(nha)	*parnu
Dyirbal	THERE	(*pangku*)	*pa*		(*paku*)
Kalkatungu	that DIST	(*payi*)	*paa*, (=*pa*)		(*pau* < *paku)

[16] For example, in a table of 3Sg pronouns and demonstratives in 16 languages that are built on the root *pa-, given in Dench (2007:228), it is clear that (a) in pronoun forms the ACC is more often distinguished (often by the suffix *-nha*) from the NOM than in demonstratives and (b) while the DAT of demonstratives is always marked by a reflex of *-ku (the normal exponent used for nouns as well), 3Sg pronoun DAT is almost invariably marked by *-ngu*, *-mpa*, or a suppletive form.

We see that it is easy to reconstruct for Proto Thura-Yura inflected forms of a stem *pa- marked with suffixes *-lu*, *-Ø*, *-nha*, and *-rnu*. The relation between *pa* and *panha* is not a straightforward distinction of case but rather a contrast between a bare form, which occurs most widely as a clitic, and an extended form (which I label ABS+ in subsequent tables). The suffix *-nha* (*-nya* in languages with a single laminal place of articulation) in many languages of the Centre and West seems to function often as a marker of specificity rather than just Accusative case. The Wirangu language, which apparently belongs to the Thura-Yura subgroup (Simpson and Hercus 2004), supplies a reflex of the expected Locative case form, *pala* 'there' beside a stem *panha* (cf. Table 7). The Adnyamathanha form *valanga* 'over there somewhere (out of sight)' cited by Alpher (2004:525) also presupposes a Thura-Yura *pala.

Dyirbal, from the North Queensland rainforest area, likewise has inflected forms of a monosyllabic root *pa-*. These function as Noun Markers and are further inflected for noun class (there is a Genitive form *pangu* not shown in Table 12). The inflections are those that are regular for nouns; however, an alternative, more common, and longer variant of the Noun Marker has the form *pala-*, which I have elsewhere interpreted as the reflex of the original Locative case form of *pa- (Koch 1995:49). The synchronic Locative of *pa-* is *palay*, which appears to involve an extension to the earlier Locative form *pala.

Note the contrast shown in Table 12 in the Dative suffix between *-ku (which in many PN languages marks the Dative of ordinary nouns) in languages where *pa- has deictic function vs. a different suffix *-rnu found on the stem when it functions as a 3Sg Personal pronoun (cf. Ngarluma *parnu-* in Table 7).

The Kalkatungu language of western Queensland has a monosyllabic distal demonstrative ('that') and definitiser with citation form *paa* (NOM), whose inflections are added to a monosyllabic stem *pa-* (Blake 1979a:34, 96). Its inflections, ERG *-yi* and Causal *-ya* in *paya*, unfortunately do not appear to be reflexes of *-lu and *-la but probably of *-thu and *tha (Blake 1979a:145–147). The main contribution of this language is that it provides further support for an inflecting monosyllabic stem *pa-. In addition to the (lengthened) Nominative form *paa* of the deictic, there is a common but content-less clitic *=pa* (Blake 1979a:96), which may be a further reflex of the suffix-less Absolutive (i.e. Nominative/Accusative) form *pa.[17]

3.2.2 Disyllabic stems continuing former inflected forms

Table 13 gives a few more examples (cf. Table 7) of disyllabic stems that are likely to continue (with no phonological changes) erstwhile inflectional forms. Dyirbal *pala* from an erstwhile Locative has already been mentioned. In both Badimaya from south-western Australia and the Pitjantjatjara and Yankunytjatjara dialects of the Western Desert language we find more than one stem, each being a reflex of a different earlier case form of a monosyllabic stem. These are examples of paradigm split — one of the kinds of morphological change mentioned in the typology given in Koch (1996). Dixon, using comparative data from Dench's (in preparation) discussion of Pilbara pronouns and demonstratives, concludes that earlier paradigms have been restructured, 'with an original case-inflected disyllabic form being taken as the new root' (Dixon 2002:335).

---

[17] In Kalkatungu's southern neighbour, Yalarnnga, *pa* likewise occurs as one of three meaningless enclitics, which are called 'prosodic suffixes' in Breen and Blake (2007:68).

Table 13: More disyllabic stems representing reflexes of inflected forms of *pa-

	palu	*panha*	*pala*
Badimaya	3Sg	that/there	
Pitj/Yank	3sg, the	that, you know the one	that (just there)
Dyirbal			THERE (Noun Marker)

The motivations for such reanalyses are not hard to discern.[18] Blake (1979b) gives many examples of the extension of Ergative and Accusative inflectional forms of personal pronouns to include the Nominative case, from which position in the paradigm they easily come to be reanalysed as the inflectional stem with no overt marker of Nominative case. The existence of ERG/NOM syncretism in many personal pronoun paradigms and of NOM/ACC syncretism in most nominals other than personal pronouns (including deictics) can easily be invoked as a factor in producing new NOM forms from erstwhile ERG or ACC forms. Dixon (1980 *passim*) has invoked a preference for disyllabic forms as a factor in replacing erstwhile monosyllabic NOM forms either by disyllabic inflected forms or by NOM forms extended by means of an incremental syllable. The reanalysis of LOC forms of deictics as stems is explained in Koch (1995), where it is shown that nominals that inherently refer to locations, i.e. deictics and place names, often occur without overt marking for Locative case. Where they happen to be provided with an overt marker, they are subject to reanalyses that treat the whole word as uninflected, with the result that the erstwhile inflection is absorbed into the stem (*A-x* is reanalysed as *Ax-Ø*) and new inflections can be added to this extended stem *(Ax-y).* The new extended stem may replace the former shorter stem in some or all of its functions; thus Dyirbal *pala-* has largely replaced *pa-* in the function of Noun Marker.

3.2.3 Particles continuing former inflected forms

Table 14 shows a few independent forms (particles) that appear to be cognate with inflected forms of *pa-.

Table 14: Particles involving *pa-*

	palu	*panha*	*pala*
Pitj/Yank	but of course	you know	look out!; indeed
Badimaya		NP coordinator	
Djapu			motion away from sp.
Gugu Nganhcara			this way, hither

In the Pitjantjatjara / Yankunytjatjara dialects of Western Desert the three particles are identical to the three disyllabic stems shown in Table 13. There is a 3Sg pronominal stem *palu-* (see Table 13), which is also used NP-finally as a marker of anaphoric status. The

[18] Dench (2007) discusses the pragmatic reasons why such paradigms split, with the former ERG form typically providing the basis for a 3Sg pronouns stem while the NOM-ACC or LOC form becomes the basis for a new demonstrative stem.

pronominal stem appears only with further inflections: ERG/NOM *-ru*, ACC *-nya*, DAT *-mpa*, LOC (stem for oblique cases) *-la*. These are regular for personal pronouns, except for the ERG/NOM *-ru*, which appears to be historically some kind of increment. One would expect that there once was an ERG/NOM bare form *palu. Such a form actually occurs in the Pintupi and Mantjiltjarra dialects (Dench 2007:228), but in Yankunytjatjara it occurs only as an uninflecting particle.

> PALU particle: 'but of course', 'only of course'. Introductory word, presenting the material that follows as established or given in some way, but also as surprising. (Goddard 1987:96).

This form apparently continues the erstwhile bare stem of the anaphoric/3Sg personal pronoun *palu-*, which was replaced in the ERG/NOM function by the extended form *paluru*.

A reflex of *panha, namely *panya*, is used both as a demonstrative stem 'that, you know the one' and as a 'presentational particle' meaning roughly 'you know'. Similarly an uninflected derivative of *panya*, namely *panyatja*, classified as a 'demonstrative adverb', is used to modify a clause with the sense 'as usual, as you know' (Goddard 1987:99).

A third reflex of *pa-, the demonstrative *pala* 'that, just there', which is 'used to shift a person's attention to something he or she hasn't been paying attention to', has a further use as an exclamatory particle 'look out!'[19] Equally idiomatic is its use in the exclamation *uwa pala!* 'yes indeed' (Goddard 1987:95).

The Badimaya Noun Phrase coordinator *panha*, which can be cliticised to either or both of a pair of co-ordinated NPs, is formally identical to the demonstrative 'that/there' mentioned in Table 13. It seems likely that this is related to the demonstrative given in Table 13, although this relationship has been stated by the author of the grammar to be 'not known' (Dunn 1988:75). The semantic development may have been 'then' > 'and'.

Djapu is one named variety of the Yolngu languages of Northeast Arnhem Land, which form a PN enclave surrounded by non-PN languages. Djapu attests a particle *bala* (< *pala: there is no voicing contrast word-initially). It is inflectionally isolated; i.e. it is not part of any inflectional paradigm. However, it does have a derivative form *balanyar(a)* 'like this' (see Table 15). The meaning of the bare form is spatial, and involves motion. It is plausibly a reflex of an earlier static locational meaning 'there', with a shift from static 'there' to a motional sense 'to there, away from here'. A second Djapu particle *bala* 'then', a co-ordinating conjunction, also appears to be a possible reflex of 'there', through the common use of locationals to indicate temporal notions — 'there' > 'then'. In another Yolngu language Djinpa the 'hither' particle *ba* is conceivably a reflex of the suffix-less stem; in many Australian languages such Absolutive forms may be used to locate a spatial zone as well as an entity; i.e. there is polysemy between 'this' and 'here', 'that' and 'there'. At any rate, *bala* appears to be independent of any inflectional paradigm. If it is a reflex of an inflected form *pa-la 'that-LOC', it would have long been deparadigmatised, judging by its particle status in a number of Yolngu languages. Yolngu forms are shown in Table 15.

---

[19] Alpher (2004:525) mentions particles *pala* and extended *palatya* in the same meaning in the Pintupi-Loritja and Warburton dialects.

Table 15: Yolngu (Northeast Arnhem Land) particle *bala*

Language	Form	Gloss	Source
Ritharrngu	*bala*	that way, thither	Heath (1980:52)
Djapu	*bala*	motion away from sp	Morphy (1983:89, 143)
Djapu	*bala*	then (co-ord conj)	Morphy (1983:139)
Djapu	*balang*	ADVERSative	Morphy (1983:144)
Djapu	*balanyar(a)*	like this	Morphy (1983:62)
Dhangu	*bala*	movement away from sp	Alpher (2004:525)
Dhangu	*bala*	and then	Alpher (2004:525)
Djinpa	*ba*	hither	Waters (1989:296)

Another language that has an isolated locational particle is Gugu Nganhcara, a Wik language from the western side of the Cape York Peninsula. Here there is a directional particle *pala*, with a clitic form *=la*, and a meaning of direction toward the speaker, 'this way, hither' (Smith and Johnson 2000:452, 442). As with Yolngu *bala*, we need to assume a change from a static to a motional sense; further we need to start from a proximal deictic sense 'this/here', which is not problematic, since the terms within deictic stems typically undergo such semantic reorganisations. Another language of Cape York, Kurtjar, has reflexes of *pala in the forms *bhal* 'that yonder', *bhalant* 'that way' (Black 1980:230). Alpher (2004:525) cites as further reflexes of *pala: a 'hither' particle from Cape York languages Wik Ngathan *pala* and Yir-Yiront *pal*, Marrgany *bala* 'that one' (called a 'demonstrative particle' in Breen 1981:346), and a Marthunira presentative particle *pala*.[20] McConvell (1996:305) mentions an element *pala* in the Ngumpin language Ngarinyman that is initial in the second of two clauses and means 'so'.

We observed in §3.1 that subordinating enclitic particles *=pa* in Nyangumarta and Ngiyampaa were treated by O'Grady as uninflected reflexes of the paradigm of pPN *pa-. Further cognates of this kind can be added. An enclitic *=pa* that marks a subordinate clause occurs in the southeastern language of the Hunter River and Lake Macquarie (Lissarrague 2006:93). In the Ngumpin languages of the northwest (to the east of Nyangumarta) a form *pa* occurs in main clauses as a 'catalyst' (also sometimes called an 'auxiliary') — with the function of indicating mood and hosting pronominal enclitics — in Mudburra and Walmajarri. An apparently related enclitic *=pa* occurs in some Ngumpin languages after clause-initial constituents (usually only if they end in a consonant).

While it is always risky to compare such short forms (which could owe their resemblance to chance), in my opinion the catalyst may have the same origin as the complementiser (cf. McConvell's claim in §3.3.3.2 for a common origin for the Gurindji complementiser and catalyst *ngu*).[21]

---

[20] '*Pala* functions as a presentative dummy taking the place of a demonstrative in a range of constructions.' (Dench 1995:185).

[21] McConvell (2006:120, pers. comm. 17.04.2009) assumes that proximate origin of the catalyst *pa* was the 'epenthetic syllable' common in a number of northwestern languages — without making any claims about a possible etymological relation with a demonstrative stem.

3.2.4 Conclusion concerning *pa-

The comparative data is summarised in Table 16. Directly attested inflectional forms of a monosyllabic root are given in bold italics. Corresponding inflecting stems are in normal italics. Cognate particles are underlined. The languages are arranged according to broad geographical regions.[22]

Table 16: Summary of evidence for inflection of *pa-

Area	Language	Gloss	ERG *palu	ABS *pa	ABS+ *panha	DAT	LOC *pala
S	p Thura-Yura	3Sg	***palu***	***pa***	***panha***	***parnu***	
S	Wirangu	that/3Sg			*panha*		*pala*
SW	Karlamay	that	***palu***				***pala***
SW	Badimaya		*palu* [a]		*panha* [c]		
W	Payungu	3Sg			*panha* [d]		*pala*
W	Thalanji	that					*pala-*
W	Ngarluma	3Sg	*palu*			***parnu-***	
W	Panyjima	3Sg			*panha*		
W	Western Desert		*palu* [b]		*panya* [e]		*pala* [f]
NW	Nyangumarta	REL		*=pa*			*pala[rri]* [g]
NW	Mudburra	COMP		*=paa*			
NW	Ngarinyman	so					*pala*
N	Djapu, Rith	that way; then					*bala*
N	Djinpa	hither		*ba*			
N	Kala Lagaw Ya	motion away		*pa*			
C	Kalkatungu	that		***paa***		***pau*** < ****paku***	
NE	Gugu Nganhcara	hither					*pala*
NE	Kurtjar	that yonder					*bhal*
NE	Dyirbal	THERE		***pa***		***paku***	*pala-*
E	Dharumbal	that			*panha* ?		
E	Marrgany	that					*bala*
SE	Ngiyampaa	SUB		*=pa*			
SE	HRLM	SUB		*=pa*			

[a] 3Sg
[b] 3Sg/the
[c] that
[d] NOM case
[e] that (Anaph)
[f] that (just there)
[g] thither

The evidence for each reconstructed form is summarised as follows. For *palu ERG, there is evidence only from languages of the south and west. For *parnu* OBL, this is likewise the case. It should be noted, however, that in these languages the stem is usually a

[22] Dench (2007:228) reconstructs, from data of this nature from 16 languages of the West, a paradigm *palu ERG, *panha NOM/ACC, *parnu DAT, *pala LOC.

personal pronoun; this may have motivated a change to its OBL inflection (from a possible *pa-ku attested in languages further east, in Dyirbal and Kalkatungu). For *panha the evidence is primarily western, plus Dharumbal of the Queensland east coast. Functionally there is no basis for reconstructing this form as a specifically ACC case form; it is more likely to have been a variant (marking specificity?) of the Absolutive *pa. For *pala LOC we have evidence form western and northwestern, Northern (Yolngu), northeastern languages (Gugu Nganchara, Kurtjar, Dyirbal), and the eastern (Maric) language Marrgany. Likewise for a zero-inflected *pa ABS we have evidence from languages of the northwest (Nyangumarta, Mudburra), south (Thura-Yura), centre (Kalkatungu), southeast (Ngiyampaa and Hunter River and Lake Macquarie (HRLM) in New South Wales), and possibly north (Djinba of the Yolngu subgroup and Kala Lagaw Ya from the West Torres Strait). I conclude that only *pa ABS and *pa-la LOC are safely reconstructable to Proto Pama-Nyungan, defined as a language ancestral to languages of the West, Centre, East, and North (Arnhem Land). Note that without consideration of the particles the evidence for *pa-la LOC would be confined to languages of a southwestern block plus Dyirbal and the evidence for *pa ABS would be confined to these plus Kalkutungu. Thus the supplementary data offered by particles has broadened the geographical spread of reflexes sufficiently to justify a firm high-level reconstruction of what looks like a partial paradigm of *pa- consisting of the ABS and LOC case forms.

3.3 Further data on *ngu-

3.3.1 Inflections of monosyllabic root

Direct evidence for the inflection of a monosyllabic stem *ngu-* is confined to a group of languages from the far west. Table 17, based largely on Dench's (2007:226) comparison of distal demonstrative forms, summarises the evidence.[23] (Note that the forms in *-nha* are never exclusively Accusative.)

Table 17: Inflected monosyllabic root *ngu-

	ERG *ngulu	NOM *ngunha	LOC *ngula	DAT/OBL *ngurnu
Ngarluma		*ngunhu*	*ngula*	*ngurna* ACC[24]
Yindjibarndi	*ngulu*	*ngunhu*	*ngula*	*ngurnu* ACC
Martuthunira	*ngulu* EFF[25]	*ngunhu*	*ngula*	*ngurnu* ACC
Thalanji	*ngulu*	*ngunha*	*ngula*	*ngurnu*
Payungu	*ngulu*	*ngunha*	*ngula*	*ngurnu*
Jiwarli	*ngulu*	*ngunha*	*ngula*	*ngurnu*
Tharrkari	*ngudu*	*ngunha*	*ngurda*	*ngurnu*

[23] Tharrkari is not in Dench's table.

[24] In some of these languages the earlier Dative has expanded to include the Accusative function, as a result of the change in alignment (cf. the following footnote).

[25] EFF stands for Effector case, whose function includes marking passive agents. It is a reflex of an earlier Ergative case. Martuthunira no longer has an Ergative case in consequence of a reorganisation of the alignment system (see Dench 1995:71).

3.3.2 Stems reflecting *ngu + inflection

Disyllabic stems involving a fixed element *ngu* followed by various increments are found especially in the Yolngu languages of Arnhem Land. The increments include the familiar suffixes *-la* and *-nha*, but also *-na* and *-ru*. These stems and their attestation are given in Table 18. In these languages deictic stems typically use a contrast of stem-final vowels to mark distal differences; e.g. *dhuwala* 'this' versus *dhuwali* 'that' in Djapu. This would appear to be the explanation for the difference between *ngula* and *nguli* in Ritharrngu and between *ngunha* and *ngunhi* in Djapu. Only one of the vowels needs to have been inherited; I assume it is the vowel *a*. It is not clear whether the apical *n* of the Djinang forms descends from a laminal *nh*. The *r* of the Oblique stem of *nguru-* is likewise unexplained historically. We can easily see, however, clear reflexes of *ngula and *ngunha, which would be cognate with the LOC and NOM forms reconstructed from the western languages in Table 17. A demonstrative with stem *ngula* also occurs in languages of northwest-central Australia, meaning 'that' in Warlpiri and 'the aforementioned' in the closely related Warlmanpa (McConvell 2006:115ff.).

Table 18: Yolngu stems involving root *ngu-*

Form	Gloss	Language
ngula-	'yon'	Ritharrngu, Djapu
nguli	Near-Distant	Ritharrngu
nguna/u-	'that'	Djinang
ngunha	'yon(der)'ABS	Djapu
ngunhi	'that/there' (anaphoric)	Djapu
nguru-	'yon' (Obl Stem)	Djapu

3.3.3 Particles based on *ngu-

Turning now to particles, we find a great number of independent words and clitics that can plausibly be understood as reflexes of inflected forms of a stem *ngu-. Table 19 gives a list, which is not exhaustive. All the examples relate to inflections for Locative or Absolutive (extended and not) cases. In the sections below I give reasons for relating these forms to the deictic stem *ngu-.

Table 19: Particles reflecting inflected forms of *ngu-

	Case	Function	Comment	Source
*ngunha	**ABS+**			
Djapu	*ngunhi*	THAT, in sub cl	= discourse deictic	Morphy (1983:127)
*ngu	**ABS**			
Kalkatungu	*nguu, ngu-*	Relative ptcl		Blake (1979a:100–103)
Ngarinyman	*ngu*	resultative complementiser		McConvell (2006:123)
Gurindji	*ngu*	catalyst		McConvell (2006:110)
Dyirbal	*ngu*	alright, that's it		Dixon (1972:124)
Djinang	*=ngu*	DEIC, pointing to ref	prob cog w *ngunung(i)*	Waters (1989:97)
Djinang	*-ngi*	Nomlsr	on Deic+Nom	Waters (1989:39)
Ritharrngu	*-ngu*	derivational suffix	e.g. *yol(ngu)*	Heath (1980:24)
*ngula	**LOC**			
Ritharrngu	*ngula*	indefinite reference		Heath (1980:50)
Djapu	*ngula*	Hypothetical ptcl	in Indic sub clause	Morphy (1983:62,127)
Djapu	*nguli*	Irrealis ptcl	conditional	Morphy (1983:62,71)
Martuthunira	*ngula*	ignorantly, I don't know	interrog	Dench (1995:166)
Pitj/Ynk	*ngula*	later, in the future		Goddard (1987:81)
Pin/Lur	*ngula*	later, by and by		Hansen and Hansen (1992:84)
pNPaman	*ngula	bye-and-bye		Hale (1976a:25)
pMPaman	*ngula	bye-and-bye		Hale (1976b:57)
Kuku Ya'u	*ngula*	now, in a while, by and by		Thompson (1988:91)
Gugu Nganhcara	*ngula*	later, afterwards		Smith and Johnson (2000:479)
Wik Mungkan	*ngul*	later on; then	v aux	Kilham et al. (1986:146, 411)
Dhangatti	*ngula*	long ago, earlier, before		Lissarrague (2007:156f)
Warlpiri	*ngula*	that, REL		McConvell (2006:115)

3.3.3.1 *ngunha

In Djapu (a Yolngu language) an uninflecting form identical to the discourse deictic *ngunhi* 'that/there' (derived by vowel substitution from *ngunha* 'yon(der)') occurs as a general clause subordinator, glossed 'THAT'. This is surely identical in origin to the deictic and this relationship demonstrates that deictics are a possible diachronic source for subordinating particles, as argued by McConvell (2006) — see below under *ngula.

3.3.3.2 *ngu

Kalkatungu has an uninflecting form *nguu* which functions as a 'relative particle'. It can host clitic pronouns =*wa* '1Sg 'and =*nha* '2Sg', in which case it has a short vowel.[26] This subordinating function of an erstwhile Absolutive form of a deictic is similar to the uses of *pa in Nyangumarta and Ngiyampaa displayed in Table 10, and to the use in subordinate clauses of the discourse deictic *ngunhi* in Djapu. A similar function is found in the resultative complementiser ('so that') *ngu* in Ngarinyman, a Ngumpin language from the Northwest (McConvell 2006:123). The same form occurs in main clauses of the Ngumpin language Gurindji as an unmarked or neutral clitic-hosting 'catalyst' or modal particle, like *pa* in Walmajarri and Mudburra (McConvell 2006:120).[27]

Another plausible reflex of Absolutive *ngu is the Dyirbal particle meaning 'alright, that's it'. If this is indeed a cognate, we need to assume that it derives from 'that/there' in a particular discourse usage. The Western Desert *uwa pala* 'yes indeed' (Goddard, 1987:95) provides a parallel; the words mean literally 'yes that' (see above).

A number of enclitics or suffixes may also derive from Absolutive *ngu, including Djinang *-ngu*.

> The DEIC affix is used to add extra deictic force. Probably the affix is historically related to the modern Distant deictic *ngunung(i*) ... the affix is found on deictics, nouns and verbs ... The most frequent use of DEIC is when a speaker wishes to point to an object or place to which he is referring, which is usually either in the field of vision of the hearer or in a certain direction. The speaker will often use DEIC in such circumstances with a pronounced rounding of the lips, which is a typical deictic gesture, in the direction of the item or place. The gesture is simultaneous with the articulation of the *-ngu* morpheme. (Waters 1989:97)

Examples show that the form is a post-inflectional enclitic. The deictic semantics, combined with the presence of a deictic root in the same language, strongly suggests that this clitic is in origin a monosyllabic, zero-inflected form of a deictic lexeme ('that, there').

Two of the Djinang demonstratives include an optional final suffix *-ngi* in the Nominative case: the IMM-PROX ('this') *djini(ngi)* and the DISTANT ('yon') *ngunu(ngi).*

> The *ngi* formative in the NOM forms ... is quite likely a reflex of the archaic -**ngu* nominaliser which occurs within such non-Djinang forms as *yolngu* 'man', 'person'. (Waters 1989:39)

---

[26] Examples are from Blake (1979a:102ff.); I owe the interpretation of the second syllables as personal pronouns to Mary Laughren (pers. comm. 28.09.2003).

[27] McConvell (2006:120) claims that the main clause usage of *ngu* is historically derivative from the subordinate usage via the process of 'insubordination', i.e. use of a subordinate clause as a main clause. There may, however, be other mechanisms for deriving both from deictics, in my opinion.

The fact that this *-ngi* is confined to the otherwise uninflected Nominative form of deictics suggests that it could be in origin an enclitic form of the Nominative of an erstwhile monosyllabic deictic stem. (The vowel change *u* > *i* is unproblematic in Djinang.) The cognate form in Ritharrngu, described as a 'derivational suffix *-ngu*', is used in the formation of certain nominal stems, such as *yol(ngu)* 'person' (Heath 1980:24). It is possibly in origin a semantically bleached and morphologically absorbed reflex of a former definitising enclitic such as Djinang *=ngu*.

The interpretation of Djinang *-ngi* as an erstwhile uninflected deictic stem (< *ngu) receives support from the behaviour of another deictic. The IMM-PROX ('this') deictic root is *dyi-*; this is evident from inflected forms *dyini* (conservative form *dyina-*) NOM, *dyirni* ERG, *dyirli* LOC. An enclitic form *=tyi(ni)* is glossed as DEF(inite) and is said to mark anaphoric definiteness, i.e. identifiability from previous reference (Waters 1989:99). The longer form appears to be identical, except for initial fortition of the stop *dy* to *ty*, with the Nominative case form of 'this'.

3.3.3.3 *ngula

We saw in the previous section that *ngula-* in Djapu and Ritharrngu is a distant deictic stem and that a 'near-distant' stem *nguli-* is derived from it by substitution of the final vowel. The same stems occur but are rare and largely restricted to the Nominative case in Ritharrngu (Heath 1980:49–50). *ngula also appears as an uninflected particle.

> At least in the Nominative, *ngula* seems to be used as a particle indicating indefinite or generalized reference ('anything', 'any kind', 'whatever')...we find the combination *nhaa ngula* 'anything (indefinite reference), including *nhaa* 'what'... (Heath 1980:50)

There appears to be a connection being made here between distant location and indefiniteness. This connection would seem to be taken further in Djapu. There the same form *ngula* occurs in non-indicative subordinate clauses as a hypothetical particle. The corresponding *i*-form *nguli* likewise functions as an irrealis particle in conditional clauses. The western language Martuthunira uses *ngula* as a particle marking ignorance after interrogatives (Dench 1995:169), in a manner reminiscent of Ritharrngu. In this language *ngula-* is otherwise the stem of the distal deictic 'that/there'.

In the Western Desert language (dialects Pitjantjatjarra, Yankunytjatjarra, Pintupi, Luritja) *ngula* is isolated from any deictic paradigm. It is a temporal particle with a meaning 'later, in the future'. The same meaning recurs in an equally isolated form in the Paman languages of Cape York Peninsula. Are these forms cognate with the deictic *ngu-*? I suggest that there is a plausible link through the temporal use of deictics, which is common in Australian languages; hence 'there (distant)' > 'then (remotely later)'. Note that 'then' is one of the glosses given for Wik Mungkan. It is possible either that the meaning 'later' developed independently in different subgroups or that it was an early development that was inherited into widely separated members of a remote high-level subgroup. In the Dhanggatti language of north-eastern New South Wales *ngula* occurs with the meaning 'long ago, earlier, before' (Lissarrague 2007:156f). This may represent a semantic shift 'there (distant)' > 'then (remotely earlier)'.

In the western language Warlpiri *ngula* is both an inflecting demonstrative and a relativising pronoun 'that' (McConvell 2006:115f). It has a particle-like usage when

followed by an enclitic particle *=tyuku* 'already'; *ngula=tyuku* in Eastern Walpiri means 'alright?' (Simpson and Nash 1990, pers. comm. 2008). A parallel form *ngula=yi* occurs in Warlmanpa with the same semantics (Nash 1979).

3.3.4 Conclusions concerning *ngu-

The evidence bearing on the reconstruction of the inflectional paradigm of *ngu is summarised in Table 20, which uses the same conventions as Table 16.

Table 20: Summary of evidence for inflection of *ngu-

Area	Language	Gloss	ERG *ngulu	ABS *ngu	ABS+ *ngunha	DAT *ngurnu	LOC *ngula
W	Thalanji	yon	***ngulu***		***ngunha***		
W	Tharrkari	yon	***ngudu***		***ngunha***	***ngurnu***	***ngurda***
W	Martuthunira	that	***ngulu***		***ngunhu***	***ngurnu***	***ngula-***
W	Western Desert	later					*ngula*
NW	Warlpiri	that, REL					*ngula*
NW	Ngarinyman	RES COMP		*ngu*			
NW	Gurindji	catalyst		*ngu*			
C	Kalkatungu	REL		*ngu(u)*			
N	Djapu	yon			*ngunha*		*ngula-*
N	Djinang	DEIC		*=ngu*			
NE	Paman	later					*ngula*
NE	Dyirbal	alright, that's it		*ngu*			
SE	Dhangatti	long ago					*ngula*

Here we see that the Locative *ngula is widely supported, with reflexes from Western, Northern, North-eastern (Paman), and South-eastern languages. The extended Absolutive *ngunha is supported by Western and Northern (Yolngu) reflexes. The bare Absolutive *ngu is sufficiently supported with reflexes from North-western, Central, Northern and North-eastern languages (if the Dyirbal form is cognate). Much more support could be amassed if the widespread incremental suffix *-ngu* could be reliably established as an absorbed reflex of a former Definite clitic. An Ergative form *ngulu and Dative/Oblique *ngurnu can only be supported from Western languages; hence these two forms are not strictly reconstructable to pPN. We do not therefore have a right to posit a pPN Ergative *ngulu on the evidence of the comparative data.[28] Nor is a Dative form reconstructable. A partial paradigm consisting of Nominative *ngu and Locative *ngula is plausibly reconstructable, but this was made possible only by the inclusion of the evidence from particles.

[28] On the other hand one might be tempted to propose a prehistoric *nguru on the basis of the 3Sg stem *nguru in Gamilaraay (SE) compared with the (oblique) stem *nguru- that occurs in Yolngu (N) and/or a *nguna based on a comparison of demonstrative stems in Yolngu (N), Warrungu, (NE) and the Yuin languages (SE). One would like, however, to have a better understanding of both the relevant historical sound changes and the possible functional force of the *-ru* and *-na* increments.

If we insist on a wide geographical distribution to justify a reconstruction as Proto Pama-Nyungan, we are forced to the conclusion that, even allowing for the inclusion of uninflected particles and inflecting stems in our comparison, we can reconstruct only Locative but not Ergative forms for two of the three CV- pronominals.[29] This is less than what can be reconstructed for another CV- pronominal stem, *nhu, as shown in Table 21. In my opinion it is likely that pPN had full paradigms, but this cannot be demonstrated from the extant forms.

Table 21: Evidence for inflectional paradigm of pPN CV pronominals

ERG	NOM	ACC	LOC	DAT/OBL
*nhulu	*nhu *ngu *pa	*nhunha	*nhula *ngula *pala	

4 Summary and conclusions

4.1 Morphological change

The data discussed here illustrates several kinds of morphological change (see Koch 1996 for a typology). Inflectional forms may be directly inherited by descendant languages. On the other hand, particular inflected forms may come to be reanalysed as stems, which in turn may take their own inflections. This reanalysis can be described in terms of the loss of a morpheme boundary, the absorption of an erstwhile affix into a lexical stem, or the demorphologisation of affixes. If the reanalysis of an erstwhile inflected form results in a form that is not capable of inflection, we have a particle.

It is to be expected, especially for pronominal words, that some case-inflected forms will end up as independent particles. This is especially true of forms that formerly expressed Locative case. It is also common for the (unmarked) Nominative case-form to become a particle. Uninflected Nominative forms, being shorter than other inflected forms, are often cliticised when they are used as particles.

4.2 Morphological reconstruction

It follows that reconstruction of the case paradigms of such pronominals can make use of the evidence that can be gained both from inflecting stems that have absorbed erstwhile inflections and from particles that have been demorphologised (or deparadigmatised or lexicalised) from erstwhile inflected members of a paradigm. Such evidence extends the available data available for reconstructing case paradigms — supplementing the data from actual inflected forms and from stems that continue reanalysed inflected forms. In fact, it is possible that in some situations it is only by using the evidence of particles that sufficient comparative data will be found to carry out a plausible reconstruction. The use of such non-morphological, basically lexical, data in performing morphological reconstruction

[29] Alpher (2004:Appendix 5.1) reconstructs *pala and *ngula as adverbs, without commenting on whether they each belonged originally to a paradigm.

constitutes an example of what I have emphasised as the 'etymological' aspect of morphological reconstruction (Koch 2003a).[30] Such an approach pursues lexical and morphological reconstruction as inter-connected activities — using the evidence of particles for reconstructing inflection (a part of grammar) and at the same time providing etymologies for independent particles (an aspect of the lexicon).

4.3 Pama-Nyungan pronominal inflection

We have seen that reliable reconstructions — based on comparative evidence and not just the projection backwards in time of inflectional patterns — can be made for the pPN Locative and Nominative inflections of *pa and *ngu only if the evidence of particles is taken into consideration. At the same time we have provided the basis for etymologising some particles. Further we have the prospect of discerning some recurrent processes of semantic/functional change involving deictics. The study could profitably be extended to other pronominal — deictic, anaphoric, and interrogative — stems. This approach holds out the prospect that — given the wider study of such phenomena, combined of course with a detailed study of changes within particular subgroups — a considerable amount of solid historical explanation will eventually be possible in the area of pronominal inflection and the etymology of particles. This in turn should aid in the clarification of the historical relations among the Australian languages. The same methodological considerations apply to other language families, where historical linguists expect the evidence of pronouns to play a large role in the demonstration of genetic relations.

References

Alpher, Barry. 2004. Pama-Nyungan: Phonological reconstruction and status as a phylogenetic group. In Bowern and Koch, eds, 2004, 93–126, 387–574, 681–687.

Anttila, Raimo. 1972. *An introduction to historical and comparative linguistics.* New York: Macmillan.

Ash, Anna, John Giacon and Amanda Lissarrague, eds. 2003. *Gamilaraay, Yuwaalaraay, and Yuwaalayaay dictionary.* Alice Springs, N.T.: IAD Press.

Austin, Peter, Corinne Williams and Stephen Wurm. 1980. The linguistic situation in north central New South Wales. In Rigsby and Sutton, eds, 1980, 167–180.

Black, Paul. 1980. Norman Paman historical phonology. In Rigsby and Sutton, eds, 1980, 181–239.

Blake, Barry J. 1979a. *A Kalkatungu grammar*. Canberra: Pacific Linguistics.

—— 1979b. Australian case systems: Some typological and historical observations. In S.A. Wurm, ed. *Australian linguistic studies*, 323–394. Canberra: Pacific Linguistics.

—— 1988. Redefining Pama-Nyungan: towards the prehistory of Australian languages. In Nicholas Evans and Steve Johnson, eds *Aboriginal Linguistics 1*, 1–90. Armidale, NSW: Department of Linguistics/University of New England.

[30] In methodological discussions such as this, Malcolm Ross is wont to comment to the effect that 'Of course one should use all the available evidence'. I have tried to make explicit here where some of that evidence is likely to be found when one is concerned with reconstructing case paradigms of pronominals.

Bowern, Claire. 2001. Karnic classification revisited. In Jane Simpson, David Nash, Mary Laughren, Peter Austin and Barry Alpher, eds *Forty years on: Ken Hale and Australian languages*, 245–261. Canberra: Pacific Linguistics.

Bowern, Claire and Harold Koch, eds. 2004. *Australian languages: classification and the comparative method.* Amsterdam/Philadelphia: John Benjamins.

Breen, J.G. 1981. Margany and Gunya. In R.M.W. Dixon and Barry J. Blake, eds *Handbook of Australian languages*, vol. 2, 274–393. Canberra: The Australian National University Press.

Breen, Gavan and Barry J. Blake. 2007. *The grammar of Yalarnnga: a language of western Queensland*. Canberra: Pacific Linguistics.

Dench, Alan Charles. 1995. *Martuthunira: a language of the Pilbara region of Western Australia*. Canberra: Pacific Linguistics.

—— 2007. Demonstrative paradigm splitting in the Pilbara languages. In Joseph C. Salmons and Shannon Dubenion-Smith, eds *Historical linguistics 2005: selected papers from the 17th International Conference on Historical Linguistics, Madison, Wisconsin, 31 July–5 August 2005*, 223–237. Amsterdam/Philadelphia: John Benjamins.

—— 2008. Getting *para* into the paradigm—an apparent example of degrammaticalization. Paper presented at the Australian Linguistic Society meeting, Sydney.

—— in preparation. Pronouns and demonstratives in the Pilbara languages of Western Australia: a comparative reconstruction. Unpublished manuscript. Perth: University of Western Australia.

Dixon, R.M.W. 1970. Proto-Australian laminals. *Oceanic Linguistics* 9.79–103.

—— 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.

—— 1980. *The languages of Australia*. Cambridge: Cambridge University Press.

—— 2002. *The Australian languages: their nature and development*. Cambridge: Cambridge University Press.

Donaldson, Tamsin. 1980. Subordination of finite sentences in Ngiyamba: (Wonga:ybuwan), a language of central western NSW. In Rigsby and Sutton, eds, 1980, 145–158.

Dunn, Leonie. 1988. Badimaya, a Western Australian language. In *Papers in Australian Linguistics No. 17*, 19–149. Canberra: Pacific Linguistics.

Evans, Bethwyn. 2003. The value of ‘applied phonology’ in morphological reconstruction. Paper presented at the annual conference of the Australian Linguistic Society, Newcastle, NSW.

Goddard, Cliff, ed. 1987. *A basic Pitjantjatjara/Yankunytjatjara to English dictionary*. Alice Springs, N.T.: Institute for Aboriginal Development.

Hale, Kenneth. 1976a. Phonological developments in particular Northern Paman languages. In Sutton, ed., 1976, 7–40.

—— 1976b. Wik reflections of Middle Paman phonology. In Sutton, ed., 1976, 50–60.

Hansen, K.C. and Hansen, L.E. 1992. *Pintupi/Luritja dictionary*. Alice Springs, N.T.: Institute for Aboriginal Development.

Heath, Jeffrey. 1980. *Basic materials in Ritharngu: grammar, texts and dictionary*. Canberra: Pacific Linguistics.

Hercus, Luise A. 1999. *A grammar of Wirangu, a language of the west coast of South Australia*. Canberra: Pacific Linguistics.

Kilham, Christine, Mabel Pamulkan, Jennifer Pootchemunka and Topsy Wolmby, eds. 1986. *Dictionary and source book of the Wik-Mungkan language*. Darwin: Summer Institute of Linguistics, Australian Aborigines Branch.

Koch, Harold. 1995. The creation of morphological zeroes. In Geert Booij and Jaap van Marle, eds *Yearbook of Morphology 1994*, 31–71. Dordrecht: Kluwer Academic.

—— 1996. Reconstruction in morphology. In Mark Durie and Malcolm Ross, eds *The comparative method reviewed: regularity and irregularity in language change*, 218–263. New York: Oxford University Press.

—— 2003a. Morphological reconstruction as an etymological method. In Barry J. Blake and Kate Burridge, eds *Historical linguistics 2001: selected papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001*, 271–291. Amsterdam/Philadelphia: John Benjamins.

—— 2003b. The case for Pama-Nyungan: evidence from inflectional morphology. In E. Hajičová, A. Kotěšovcová and J. Mírovský, eds *Proceedings of XVII International Congress of Linguists, Prague, Czech Republic, July 24–29, 2003.* Prague: Matfyzpress, Matematicko-Fyzikální Fakulty, Univerzity Karlovi. [CD-ROM]

Lissarrague, Amanda. 2006. *A salvage grammar and wordlist of the language from the Hunter River and Lake Macquarie*. Nambucca Heads, NSW: Muurrbay Aboriginal Language and Culture Cooperative.

—— 2007. *Dhangatti grammar and dictionary with Dhangatti stories*. Nambucca Heads, NSW: Muurrbay Aboriginal Language and Culture Cooperative.

Mathews, R.H. 1904. The Ngunawal language. *Journal of the Royal Anthropological Society* 34.294–299.

Mathews, R.H. and M.M. Everitt. 1900. The organisation, language, and initiation ceremonies of the Aborigines of the south-east coast of New South Wales. *Journal of the Royal Society of New South Wales* 34.262–281.

Meillet, Antoine. 1967. *The comparative method in historical linguistics*. Translated by Gordon B. Ford, Jr. (Instituttet for Sammenlignende Kulturforskning) Paris: Champion.

McConvell, Patrick. 1996. The functions of split-Wackernagel clitic systems: pronominal clitics in the Ngumpin languages (Pama-Nyungan family, Northern Australia). In Aaron L. Halpern and Arnold M. Zwicky, eds *Approaching second: second-position clitics and related phenomena*, 299–331. Stanford: CSLI.

—— 2006. Grammaticalization of demonstratives as subordinate complementizers in Ngumpin-Yapa. *Australian Journal of Linguistics* 26.107–137.

Morphy, Frances. 1983. Djapu, a Yolngu dialect. In R.M.W. Dixon and Barry J. Blake, eds *Handbook of Australian languages*, vol. 3, 1–188. Canberra: The Australian National University Press.

Nash, David. 1979. Preliminary vocabulary of the Warlmanpa language. Unpublished manuscript. Cambridge, Mass.: Massachusetts Institute of Technology.

O'Grady, Geoffrey N. 1990. Pama-Nyungan: the tip of the lexical iceberg. In G.N. O'Grady and D.T. Tryon, eds *Studies in comparative Pama-Nyungan*, 209–259. Canberra: Pacific Linguistics.

Rigsby, B. and P. Sutton, eds. 1980. *Papers in Australian linguistics No. 13: contributions to Australian linguistics*. Canberra: Pacific Linguistics.

Ross, Malcolm. 2001. Is there an East Papuan phylum? Evidence from pronouns. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

—— 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew Pawley, Robert Attenborough, Jack Golson and Robin Hide, eds *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Pacific Linguistics.

Sands, Kristina. 1996. *The ergative in Proto-Australian*. München: Lincom Europa.

Simpson, Jane and David Nash. 1990. Wakirti Warlpiri. Unpublished manuscript. Canberra: AIATSIS.

Simpson, Jane and Luise Hercus. 2004. Thura-Yura as a subgroup. In Bowern and Koch, eds, 2004, 179-206.

Smith, Ian and Steve Johnson. 2000. Kugu Nganhcara. In R.M.W. Dixon and Barry J. Blake, eds *Handbook of Australian languages*, vol. 5, 355–507. Melbourne: Oxford University Press.

Sutton, Peter, ed. 1976. *Languages of Cape York*. Canberra: Australian Institute of Aboriginal Studies.

Thompson, David. 1988. *Lockhardt River 'sand beach' language: an outline of Kuuku Ya'u and Umpila*. (Work Papers of SIL-AAIB, A11) Darwin: Summer Institute of Linguistics.

Trask, R.L. 2000. *The dictionary of historical and comparative linguistics*. Edinburgh: Edinburgh University Press.

Tsunoda, Tasaku. 2003. *A provisional Warrungu dictionary*. (ICHEL Linguistic Series vol. 8) Tokyo: Department of Asian and Pacific Linguistics, University of Tokyo.

Waters, Bruce E. 1989. *Djinang and Djinpa: a grammatical and historical perspective*: Canberra: Pacific Linguistics.

Williams, Corinne. 1980. *A grammar of Yuwaalaraay*. Canberra: Pacific Linguistics.

Wurm, S.A. 1972. *Languages of Australia and Tasmania*. The Hague: Mouton.

18 *Yet more Proto Austronesian infixes*

PAUL JEN-KUEI LI AND SHIGERU TSUCHIDA

1 Introduction[1]

Compared to prefixation and suffixation, infixation is a rare morphological process in language. Yet it is commonly found in Austronesian languages.

Two infixes, *-um- and *-in-, have been reconstructed for Proto Austronesian (Wolff 1973; Ross 1995, 2009). They are productive and widespread in Formosan and western Austronesian languages. So their reconstruction is generally accepted.

Dempwolff (1934–38) reconstructs some lexical forms containing nonproductive infixes called 'erstarrte Infixe' (petrified infixes), including *-al-, *-ar- and *-aR-, as in *p<al>apah 'plant material', *p<al>upuh 'plant material', *p<aR>atpat 'vegetable matter', without further comment. Lopez (1977) discusses this problem in some detail, lists many examples in western Austronesian languages, including Malay, Javanese, Ngadju-Dayak, Toba-Batak, and several Philippine languages, but with only a few examples from Formosan languages, and proposes to add *-el- and *-er- to Dempwolff's *-al-, *-ar- and *-aR-. Blust (2009:378) also states that 'there is widespread support for infixes *-ar-, *-al- or *-aR-'. He cites only a few uncertain examples from Pazih and Paiwan, but many examples from several western Austronesian languages, including Hanunoo, Central Tagbanwa, Malay, Toba Batak and Sundanese. In this paper we shall reassess the reconstruction of these nonproductive infixes by providing as many examples from Formosan languages as possible and some examples from extra-Formosan languages, such as Tagalog and Javanese. While we agree with Dempwolff, Lopez and Blust in reconstructing *-al- and *-aR-, we have some reservation for *-ar-, which is attested in only one Formosan language (Paiwan) and some western Austronesian languages.

[1] Our friend Malcolm Ross has made great contributions to Proto Austronesian reconstructions over these years. This paper may only serve as a footnote to his illuminating papers. It was drafted by the first author. Shigeru Tsuchida provided the first author with many examples from several Formosan languages, including Tsou, Kanakanavu, Saaroa, Bunun, Paiwan, Puyuma, Rukai, Siraya, and Amis, as well as Tagalog and Makassarese. We have benefited from Lopez's (1977) earlier study and comments from Robert Blust, Bethwyn Evans, Andrew Pawley, Elizabeth Zeitoun, and two anonymous reviewers. This work was supported in part with a grant from the National Science Council (NSC 95-2411-H-001-010-H).

Moreover, we propose to reconstruct *-aN-, whose reflexes are found in most Formosan languages and in at least two western Austronesian languages, Tagalog and Makassarese.

These nonproductive infixes appear mostly in fossilised forms attested in most, if not all, Formosan languages that belong to different major subgroups, as based on Blust's (1999a) subgrouping hypothesis,[2] including (1) Paiwan, (2) Puyuma, (3) Bunun, (4) Tsou, Kanakanavu and Saaroa of the so-called Tsouic subgroup, (5) Kavalan, Basay, Amis and Siraya, which belong to East Formosan, (6) Thao and Favorlang, which belong to Western Plains, (7) Saisiyat and Pazih, which belong to Northwest Formosan, (8) Atayalic, and (9) Rukai. These languages are geographically far apart in Taiwan, from the north (Basay, Atayal and Kavalan) to the south (Paiwan and Siraya) and from the west (Pazih and Thao) to the east (Amis and Puyuma). In fact, these infixes are also attested in extra-Formosan languages, such as Tagalog and Javanese. Thus diffusion may not be a plausible explanation for their existence. Parallel development is also unlikely.

Lexical forms with the infixes under discussion are numerous in Paiwan, but limited in most other Formosan and western Austronesian languages. Some of these infixes seem to be partially productive in Paiwan as well as Thao and Amis, but completely nonproductive and fossilised in all other languages. However, when we tested a few lexical forms with or without these infixes, our Paiwan informant was not aware of their relationship.

Like the commonly found infixes, *-um-* and *-in-*, in Austronesian languages, the uncommon infixes are usually inserted after the initial consonant of the root and they are usually in the form of *-aC-*. However, they may appear initially if the root begins with a vowel, e.g. *alh-eses* 'a tiny sound' in Amis. In fact, these nonproductive infixes also appear as prefixes, regardless of the root shape, in some Austronesian languages, just as do *-um-* and *-in-*.

2 Infixes in mostly fossilised forms in Formosan and extra-Formosan languages

Examples are listed below as exhaustively as possible in Formosan languages, except Paiwan, which has an unusually large number of lexical forms with these uncommon infixes.

2.1 Thao and Favorlang

In addition to the infixes, *-um-* and *-in-* or their variants, Thao has three more infixes, *-ar-*, *-al-* and *-az-*.[3]

---

[2] Blust's (1999) classification of the Formosan languages, based on shared innovations in phonology, is as follows: (1) Atayalic, (2) East Formosan, (3) Puyuma, (4) Paiwan, (5) Rukai, (6) Tsouic, (7) Bunun, (8) Western Plains, and (9) Northwest Formosan. In his view, these nine Formosan subgroups plus Malayo-Polynesian constitute all the primary branches of PAn. Malcolm Ross (2009) proposed a modified subgrouping hypothesis, in which there are four major branches, Tsou, Rukai, Puyuma, and Nuclear-Austronesian. If their subgrouping hypotheses are correct, then the three special infixes under discussion occur in all the primary branches of Austronesian.

[3] A fourth infix *-ur-* is found in the form *b<ur>uqbuq* 'bubble' (Blust 2003:335). But the form of the infix is not the *-aC-* under discussion although it may have undergone assimilation: *-ur-* < *-ar-*.

Examples of *-ar-*: *q<ar>afqaf* 'house (arch.)', *b<ar>imbin*[4] 'vehicle', *b<ar>umbun* 'rumbling sound of thunder', *t<m>u-b<ar>umbun* 'rumble', *f<ar>uzfuz* 'dandruff', *k<ar>ungkun* 'to wind', *sh<ar>inshin* 'bell', *p<ar>akpak* 'make a popping sound' < *pakpak* 'clap' (Blust 2003:95), *k<ar>ashkash* 'sound of crashing in the underbrush' (Blust 2003:449) < *kashkash* 'scratch up' (Blust 2003:454), *ma-'<ar>uz'uz* 'lively', *ya-'<ar>uz'uz* 'vigorous, in good spirit', *ma-k<ar>ishkish* 'diligent', *ma-f<ar>ushfush* 'rough', *mak-ma-m<ar>ismis* 'keep blinking' < *mismis* 'eyelash, blink' (Blust 2003:95), *mashi-k<ar>uskus* 'coil, curl', *mashi-k<ar>upkup* 'tangled, as thread' (Blust 2003:592), *mashi-k<ar>uthkuth* 'wrinkled', *makit-ha-h<ar>umhum* 'chatter interminably', *mak-b<ar>uqbuq* 'clear one's throat, gush out', *b<ar>amban* 'plant sp.', *f<ar>izfiz* 'bamboo leaves', *ma-h<ar>aqhaq* 'loose', *k<ar>itkit* 'a curl', *ma-qa-q<ar>izqiz* 'noisy', *ma-sa-s<ar>psap* 'talkative, gossipy', *ma-t<ar>ithtith* 'blunt, dull'.

Examples of *-al-*: *b<al>umbun* 'small ringing bell for dancing', *sh<al>inshin* 'bell', *q<m>ay-sh<al>inshin* 'ring a bell', *b<al>akbak* 'tap' (Blust 2003:316).

Examples of *-az-*: *ma-h<az>iwhiw* 'cool', *k<az>inkin* 'earring', *lhin-da-d<az>umdum* 'cut into small pieces', *ma-p<az>ashpash* 'soft, loose'.

We have found more than twenty examples for the infix *-ar-*, but only a few for *-al-* in Thao, as given above. The former is historically derived from PAn *-al-, while the origin of the latter is unknown. The original source of /l/ in Thao may have been from Bunun /l/ (< *R) due to close contact. However, it is not clear whether the source of the Thao forms with *-al-* can still be found in Bunun, as unfortunately there is no extensive and reliable Bunun dictionary available yet. Notice that there is a clear contrast between *-ar-* and *-al-* in Thao, as in *b<ar>umbun* 'rumbling sound of thunder' versus *b<al>umbun* 'small ringing bell for dancing'. Thao *-az-* is historically derived from *-aN-, *-an'- or *-aj-.

We have also found a few examples of the Favorlang infix *-ar-*, which reflects PAn *-al-, e.g., *r<ar>icherich* 'a strip marked off with a knife for a measure', *s<arr>oso* 'mist, damp', cf. *soso* 'length', and the infix *-en-*, e.g., *l<en>esoleso* 'shout after the native manner' (Happart 1650, cited in Campbell 1896).

It does not seem to be accidental that two closely related languages, such as Thao and Favorlang, share the same reflex, *-ar-* < *-al-.

2.2 Kavalan and Basay

2.2.1 Kavalan

Kavalan has four nonproductive infixes, *-(a)r-*,[5] *-(a)R-*, *-(a)n-*, and *-l-*:

Examples of *-ar-* or *-r-*: *k<ar>makmaz* 'blink', cf. *kmaz* 'eyelash', *t<r>aqitaq* 'talkative', *p<r>inipin* 'walk unsteadily', *k<r>awkaway* 'work', *qabaw* or *k'<r>avao* [sic] 'buffalo' (Taintor 1874), *k<r>amkamut* 'fast', *k<r>ibakib* 'a type of fern, Asplenium nidus', *k<r>imkimay* 'a type of coral', *q<r>imqimun* 'open and close', *q<r>iwqiw* 'wag (tail)'.

---

[4] Note the homorganic assimilation of the nasal to the following stop in word-medial position in some of these forms.

[5] There is a general vowel-deletion rule, which also applies to the infixes, *-um-, -in-, -ar-,* and *-aR-* in Kavalan.

Examples of *-aR-* or *-R-*:[6] *t<R>abtab* 'eat and make noise like a pig', *b<aR>qian* 'old people' < *baqi* 'grandpa', *q<R>ezqez = qezqez* 'stable', *q<R>itun = qitun* 'vehicle', *q<R>ibi = qibi* 'large container', *q<R>utay = qutay* 'plant sp., Diplazium esculentum Sw.' *q<R>apqap = qapqap,* 'grope in the dark', *q<R>ungqung* 'noise'.

Examples of *-an-* or *-n-*: *b<n>aqi-an* 'elder' < *baqi* 'grandfather', *b<n>aRqi-an* 'old people' < *baRqi-an* 'old people' < *baqi* 'grandfather', *s<an>ayasay = s<n>ayasay* 'Green Island', *t<n>ubtub* 'plant sp., Macaranga tanarius (L.)'.

Examples of *-l-*: *p<l>iapia* 'plant sp., Marislea crenata Presl'.

Both Kavalan /r/ and /R/ are historically derived from *l or *R, although the conditions are unknown; /n/ is derived from *n, *n', *j or *N, and /l/ from *R (Li and Tsuchida 2006:8–9).

2.2.2 Basay

Basay has three nonproductive infixes, *-al-, -ar-,* and *-an-*, as in (Asai 1991):

Examples of *-al-*: *b<al>akibaki* 'old man', cf. *baki* 'grandfather', *h<al>iuhiu* 'assistant', *k<al>imikimi* 'wink', *l<al>isilisi* 'languid, weary', *p<al>angpang* 'pot, water jar', cf. *pangpang* 'pond, lake', *t<al>aktak = taktak* 'cut, chop'.

Example of *-ar-*: *s<ar>abaisabai* 'haunt'.

Examples of *-an-*: *k<an>engkeng* 'lapis lazuli', *k<an>aba* 'mortar', cf. *kaba* 'clothing'.

Basay /l/ or /r/ is historically derived from *R or *D, and /n/ from *n, *N or *j.

2.3 Siraya

There are two nonproductive infixes, *-al-* and *-ar-*, in Siraya (Adelaar 2000:36; Murakami 1933). They appear in monosyllabic root reduplicated stems, mostly in fossilised forms, as in:

Examples of *-ar-:* *h<ar>ilhil* 'smooth', *k<ar>otkot* 'small river', *s<ar>amsam* 'useless', *v<ar>igbig* 'bore, drill', *va-v<ar>ingbing* 'run violently down', *mi-h<ar>afhaf* 'troubled, confused', *paka-p<ar>upu* 'crush, grind'.

Examples of *-al-:* *p<al>i(x)pix* 'fine, refined' < *pixpix* 'crumbs', *p<al>ungpung* 'cease (of wind)', *k<al>ongkong* 'nail', *t<al>octock* 'hat' (cf. *toucktouck* 'crown, top').

2.4 Amis

There are five nonproductive infixes, *-al-, -ar-*, *-an-*, *-alh-* and *-a'-* in Amis with quite a few examples for the first two, as in (Fey 1986; Zeng 2008)[7]:

---

[6] Some of these examples contain a uvular stop *q* in the stems, and so the infix *-aR-* or *-R-* may have originally been *-ar-* or *-r-* (< *l), with *-(a)R-* resulting from assimilation to the adjacent *q*.

[7] We have slightly modified some of Fey's orthography in that we use *ng* instead of *g* to stand for a velar nasal and *lh* instead of *d* to stand for a voiceless lateral in the prestige dialect of Amis. Similarly, we use *lh* instead of *d* and *o* instead of *u* in Zeng's orthography.

Examples of *-al-*: *ng<al>iwngiw* ‘complain’, cf. *pa-ngiwngiw* ‘speak ill behind one’s back’, *c<al>ascas* ‘sound of running water’, *c<al>emcem* ‘fear’, *c<al>aycay* ‘bear lots of fruit’, *c<al>epcep* ‘cautious’, *c<al>ikcik* ‘itchy’, *f<al>angfang = f<al>ungfung* ‘right size (clothes)’, *f<al>atfat* ‘twinkle’, *f<al>itfit* ‘prickly heat’, *f<al>iwfiw* ‘sound of a gentle breeze’ < *fiwfiw* ‘tiny sound’, *f<al>awfaw-an* ‘north wind’, *f<al>okfok* ‘sound of water or clapping of wings’, *f<al>ongfong* ‘loose’, *ng<al>iwngiw* ‘complain’, *ng<al>o'ngo'* ‘the sound of voices’, *h<al>amham* ‘deep expectancy’ < *hamham* ‘expect something’, *k<al>amkam* ‘diligent’, *k<al>angkang* ‘ringing sound’, *k<al>askas* ‘noise of steps’, *k<al>atkat* ‘glisten’, *k<al>eskes* ‘unpleasant’, *k<al>ilhkilh* ‘rapid flow of water’, cf. *kilhkilh* ‘drag something’, *k<al>iling* ‘spoon’, cf. Paiwan *kizing, k<al>ongkong* ‘echo of thunder’, cf. *kongkong* ‘knock at a door or drum’, *k<al>ongkong-ay* ‘green pepper’, *k<al>ohkoh* ‘noise of walking in rain’, *p<al>okpok* ‘grass floating on a pond’, *s<al>angsang = s<al>a'sa'* ‘sound of rain or waterfall, noisy’, cf. *sa'sa'* ‘drain through a sieve’, *s<al>ingsing* ‘sounds of bells ringing’, *t<al>afaelh = tafaelh* ‘go above’, *t<al>a'ta'* ‘sound of rain’, *t<al>aktak* ‘dripping of water’, cf. *taktak* ‘pour out of a container’, *t<al>engteng* ‘throb with pain’, *w<al>alhwalh = w<al>a'wa'* ‘nauseated’.

Examples of *-ar-*: *c<ar>ofacof* ‘fog, light cloud’, *c<ar>angcang* ‘smoothly’, *c<ar>ingcing* ‘loud and clear’, *c<ar>ofacof* ‘spring’, *f<ar>awafaw* ‘mild weather’, *f<ar>axfax* ‘impetuous’, *f<ar>itfit* ‘have rash’, *h<ar>okhokay* ‘large intestines’, *h<ar>engheng* ‘the sound of waves’ < *hengheng* ‘rough, heavy sea’, *k<ar>askas* ‘noise of wind’, *k<ar>engkeng* ‘noise of a landslide’, *k<ar>ingking* ‘striking sound’, *k<ar>a'ka'* ‘the sound of a creaky wagon’ < *ka'ka'* ‘roar with laughter’, *k<ar>ihkih* ‘rustling sound’, *k<ar>iwkiw(-ay)* ‘hermit crab’, *ng<ar>awngaw* ‘the noise of many voices’, cf. *ngawngaw* ‘eat something raw’, *p<ar>ekpek* ‘march’, *p<ar>okpok* ‘gallop’, *r<ar>awraw = rawraw* ‘noisy’, *s<ar>insin* ‘dizzy’, *s<ar>iwsiw* ‘breeze’, *t<ar>amtam* ‘spicy’, *t<ar>astas* ‘noise of a rattan bed’, *t<ar>ingting* ‘firm and stand straight’, *t<ar>iktik = t<an>iktik* ‘walk lightly and gracefully’, *t<ar>nes/tnes* ‘delay’, *t<ar>angitang* ‘rose-colored’, *t<ar>oktok* ‘palpitate’.

Examples of *-an-*: *c<an>iwciw* ‘throes’, *f<an>ekfek* ‘skilful’, *f<an>etfet* ‘hot and suffocating’, *h<an>anghang* ‘pain’, *h<an>inghing* ‘smell of stool’, *k<an>awkaw* ‘enthusiastic’, *s<an>alsal* ‘tickle’, *s<an>awsaw* ‘unsettled in spirit’, *s<an>engseng* ‘just right’, *t<an>ektek* ‘firm, solid’, *t<an>estes* ‘tidy’.

Examples of *-alh-*: *alh-eses* ‘tiny sound’, *c<alh>amcam* ‘same’, *c<alh>ofacof* ‘mist, fog’, *c<alh>opcop* ‘crowded’, *c<alh>i'ci'* ‘hurt, sting of a wound’, *h<alh>okhok* ‘cry’, *lh<alh>eclhec* ‘dense’, *lh<alh>iclhic* ‘full’, *p<alh>akopak* ‘paper’, *p<alh>alhoki = p<elh>alhoki* ‘bat’ < *p<aN>iki,[8] *p<alh>ongipong* ‘type of flower, Common Melastoma’, *s<alh>amsam*’spicy, hot (taste)’, *s<alh>ingsing* ‘cool (weather)’, *s<alh>iwasiw* ‘sharp, pointed’, *t<ar>olho'* ‘fingers, toes’ by metathesis < *t<alh>oro' < *tuZuq ‘point at’.

Example of *-a'-* : *t<a'>sel/tsel* ‘pierce through’.

Amis derivations from PAn are as follows: *l* < *l, *R, *r* < *D, *n* < *n, *j, *lh* < *N, and ' < *q.

---

[8] Notice that there is an unanticipated vowel /o/ instead of /i/ in the Amis forms.

2.5 Saisiyat and Pazih

2.5.1 Saisiyat

Saisiyat has two main nonproductive infixes, *-aL-* ~ *-æL-*[9] and *-al-* ~ *-l-*, as in (Tsuchida 1962; Li 1978):

Examples of *-aL* ~ *-æL-*: *k<aL>izikiz* 'side of armpit', *s<aL>abong* 'gather moss due to lack of the sun', cf. *sabong* 'shade', *'<æL>ez'ez* 'bitter'.

Examples of *-al-* ~ *-l-*: *k<al>ongkong-an* 'very skinny', *ka-k<l>okoeh* 'fingernail' < *kuS_2kuS_2 (Dahl 1981:35)(cf. *k<alj>uskus-an* 'fingernail' in Paiwan).

Saisiyat /L/ is historically derived from *l or *R and /l/ from *N or *n'. Saisiyat /L/, which is a flap, was retained only in the speech of a few older speakers of the Taai dialect in 1970s.

2.5.2 Pazih

There are three nonproductive infixes, *-al-, -ar-* and *-a-*, in Pazih, as in (Li and Tsuchida 2001:18–19):

Example of *-al-*: *k<al>ikux* 'fingernail' < *kuS_2kuS_2 (Lopez 1977. Note the irregular reflex of the first vowel *i* and that of the final consonant *x* of the root).

Examples of *-ar-*: *b<ar>anaban* 'urn', *dungudung ~ d<ar>ungudung* 'drum'

Examples of *-a-*: s*<a>ungusung* 'is counting', *p<a>isuzuk* 'is hiding'.

Pazih /l/ is historically derived from PAn *N or *n', and /r/ from *l or possibly *r.[10] The above examples for both derivations, *-al-* < *-aN- and *-ar-* < *-al- or *-ar-, are problematic.

Pazih is one of four Formosan languages which has an infix with the vowel /a/ without a following consonant in the first syllable of the stem. Another language is Bunun, which has an infix *-a-* in *l<a>ihlih* 'car.' PAn *l is lost in Bunun, so it may reflect *-al-. The third language is Paiwan, which also seems to have the infix *-a-* found in two problematic frozen forms, *dj<a>ungadjing-an* 'cock's comb', *m<a>udmid* 'dry and brittle'[11] (Ferrell 1982:16). However, PAn *R is lost in Paiwan, and so the Paiwan infix *-a-* may reflect *-aR-. The fourth language is Puyuma, as in *tu<a>ktuk* 'hammer'. However, here *-a-* is inserted between the first vowel and the following consonant, rather than immediately after the initial consonant of the stem as found in all other languages. It is thus a different type of infix.

2.6 Bunun

There are four nonproductive infixes, *-an-*, *-al-*, *-az-*, and *-a-*, in Bunun, as in (Nihira 1988 and Nojima, pers. comm.):

---

[9] Saisiyat /L/ was retained only in a few older speakers of the Taai dialect (Li 1978), and examples containing it are based on that dialect. In Saisiyat *a split into two different vowels, /a/ and /æ/, when adjacent to the glottal /'/ or /h/.

[10] In Blust's (1999b:333–334) study, he found 34 lexical items containing /r/ which is derived from *l, but only three items containing /r/ which is derived from *r.

[11] Note that the two vowels, *u* and *i*, in each root are not identical, so we cannot be certain whether *a* in these forms is an infix in Paiwan.

Examples of *-an-*: *t<an>uduq* 'finger' < **tuZuq, pis-t<an>aqtaq* 'play', *b<an>itul* 'plant sp., Dendrobium flaviflorum' < **bituR,*[12] *ma-s<an>ingsing* 'clean'.

Examples of *-al-*: *v<al>ungvung* 'bell', *k<al>angkang* 'sow, female pig or wildboar', *s<al>avsav-az* 'plant sp., Panicum plictum', *s<al>uksuk-az* 'plant sp. (with thorny stems)', cf. *suksuk* 'nape', *t<al>untun-az* 'shaft of a spear', *t<al>ungtung* 'a unit of quantity'.

Examples of *-az-*: *k<az>ingking* 'hang (transitive)', cf. *mal-k<az>ingking* 'hanging down', *an-k<az>ingking* 'carry something by hanging', *ma-p<az>aspas* '(cooked rice) not sticky and dropping from the mouth', cf. *mu-paspas* '(leaves, fruits) drop, fall', *p<az>ikpik* 'always tell a lie'.

Example of *-a-*: *l<a>ihlih* or *laihlaih* 'car' (uncertain example).

Bunun /n/ is derived from *n, *n' or *N, /l/ from *R, and /z/ from *y. PAn *l is lost, so Bunun infix *-a-* may reflect *-al-.

2.7 Paiwan

The infixes of <*ar*>, <*al*> and <*alj*> occur in quite a few lexical forms in Paiwan, often with the meaning 'having sound or quality of' (Ferrell 1982:16), and they occur right after a wide range of initial consonants. However, we found only one example for *-a-* and a few problematic examples for *-el-* and *-er-*.

Examples of *-ar-*: *dj<ar>emdjem* 'fine mist falls', cf. *djemdjem* 'push down', *k<ar>uDkuD* 'have sound of hoeing' < *kuDkuD* 'hoe', *k<ar>uskus* 'have sound of scraping' < *kuskus* 'scraper', *s<ar>apsap* 'grope through' < *sapsap* 'scratch in earth', *s<ar>igsig* 'crackle (something dry)', *b<ar>engbeng* 'make buzzing sound', *ts<ar>abtsab* 'make slapping noise', *g<ar>avagav* 'crawl, wriggle', *b<ar>uqbuq* 'make sound of water boiling', cf. *buqbuq* 'push someone into water', *D<ar>apDap* 'feel around in mid-air', *D<ar>epDep* 'glass beads', *D<ar>iwDiw* 'signal with burning brand' < *DiwDiw* 'burning brand used as a signal', *g<ar>aljigalj* 'bed or sleeping shelf', *k<ar>abkab* 'make sound of flapping', *p<ar>etspets* 'flap wings', cf. *mi-petspets* 'flap wings', *q<ar>awqaw* 'occur sound of shouting', cf. *q<m>awaaw* 'shout loudly', *t<ar>ivtiv* 'plant sp.', *tj<ar>ra-keDi* 'definitely small' < *tja-keDi* 'smaller' < *keDi* 'small'.

Examples of *-al-*: *g<al>emgem* 'furious' < *gemgem* 'fist', *k<al>edjip* 'wink' < *kedjip* 'eyelash', *s<al>apsap* 'scratch in earth' < *sapsap* 'scratch in earth', *z<al>angzang* 'perspire' < *zangzang* 'body heat', *v<al>eljvelj* 'pretty', cf. *veljvelj* 'banana', *dj<al>awdjaw* 'placenta', *djawdjaw* 'leaves of sweet potato plant', *D<al>emDem* 'light green', *g<al>iljgilj* 'shining', *ng<al>asngas* 'pant', *p<al>aqpaq* 'astringent', *q<al>amqam* 'smooth, slippery', *t<ar>aktak* 'snap open', *tj<al>aDtjaD* 'jerky, jolting', *ts<al>eqtseq* 'make sound of a falling tree', *v<al>ljvalj* 'nauseated', *z<al>engzeng* 'make sound of wind'.

---

[12] Compare Tsou (Duhtu) *fitru* 'id', Kanakanavu *v<an>turu*, Saaroa *v<alh>ituru* 'sp. of orchid, Dendrobium flaviflorum'.

Examples of *-alj-*: *b<alj>angbang* 'make sound like rain on tin roof', *dj<alj>emdjem* 'give help', *g<alj>awgav-an* 'palm of hand, fingers', cf. *gawgaw* 'accept, receive', *k<alj>engkeng* 'have ringing in ears', *k<alj>edjip* 'blink eye' < *kedjip* 'eyelash', *k<alj>uskus-an* 'fingernail' < *kuskus* 'scraper', *p<alj>uDpuD* 'heart beats fast', *q<alj>emqem* 'sweet-taste', cf. *qemqem* 'mouth full', *s<alj>emsem* 'have throbbing ache', *ts<alj>iptsip* 'make clicking noise in mouth', *tj<alj>a-keDi-keDi* 'smallest' < *tja-keDi* 'smaller' < *keDi* 'small'.

Example of *-a-*: *s<a>ungusung* 'plant sp., Debregeasia edlis, whose berries edible'.

Example of *-el-*: *ts<el>alaq* 'thunderclap', cf. *tselaq* 'a crack or split'.

Examples of *-er-*: *dj<m><er>akuts* 'pick up with talons', cf. *djakuts* 'grab', *se-k<er>umulj* 'fall flat', cf. *tja-kumulj* 'quilt, coverlet'.

Notice that *-el-* and *-er-* appear only in antepenultimate position.

Paiwan derivations from PAn are: l < *l, r < *r, lj < *N and ø < *R. It is not clear whether *-a-* in Paiwan reflects *-aR- or *-a-.

2.8 Puyuma

The infix *-al-* in Puyuma 'attaches to nominal and verbal bases and means "having the sound of"...,' as in Cauquelin (2008:14):

Examples of *-al-*: *s-al-iwsiw* 'twittering of spring chickens' < *siwsiw* 'spring chicken', *p-al-etik* 'sound of splashing water, of popping rice' < *petik* 'splash, pop', *s-al-teb* 'sound of a chopping knife on a wooden board' < *tebteb* 'chopping knife', *s-al-ti*k 'crackling sound of matches' < *tiktik* 'tattoo'.[13]

According to Teng (2008:36–37),[14] there are four infixes of this type, *-al-, -aL-, -ar-,* and *-ag-*, and her examples are listed below. Both Cauquelin and Teng worked on the Nanwang dialect of Puyuma, while Tsuchida's (1980, 1983) work is based on the Tamalakaw dialect of Puyuma. His examples are listed after Teng's below. He found these four infixes: *-al-*, *-aL-*, *-ar-*, and *-a-*.

Examples of *-al-*: *g<al>emgem* 'numbness of the tongue', *b<al>eTbeT* 'recurring pains', *t<al>ustus* 'prickle', *s<al>engseng* 'lonely', *T<al>ebteb* 'nervous' (Teng 2008:37); *T<al>ingTing* 'cold (water, rain)', *H<al>angaHang-an* 'bark (as a dog)', cf. HangHang 'bark (as a dog)', *pa-t<al>ustus* 'cry (of a rat)', *pa-T<al>unTun* 'make a noise of knocking at a door', *k<al>ungkung* 'empty', *h<al>emhem* 'refreshing taste', *s<al>i<a>HsiH* 'sharp pain', cf. *siHsiH* 'sprinkle water' (Tsuchida 1983).

---

[13] As noted in Cauquelin (2008:14), 'when the initial consonant is /t/ in the onomatopoeic, it becomes /s/', as in the last two examples.

[14] We have interpreted certain forms differently from Teng. She analyses the form *baLangabang* 'easy clothes' as containing two infixes *b<aL>ang<a>bang*, whereas we would analyse it as containing only the first infix *-aL-* and treat the second *-a-* as an empty vowel conjoining two identical syllables. Similarly, the vowel *-e-* (schwa) in the form *T<ag>ageTag* 'pour out' can be left out.

Examples of *-aL-*: *b<aL>angabang* 'easy clothes', *b<aL>ukbuk* 'easy shoes or pants', *s<aL>iksik* 'high-spirited' (Teng 2008:37)*; v<aL>angavang* 'loose (clothes)', *k<aL>angkang* 'empty', *pa-k<aL>angkang* 'make a noise', *pa-k<aL>i<a>ngking* 'clatter' (Tsuchida 1983).

Examples of *-ar-*: *s<ar>ibsib* 'smooth skin', *T<ar>isTis* 'noisy' (Teng 2008:37)*; pa-s<ar>ksak* 'make a rustling noise', *pa-s<ar>ingsing* 'make a clinking sound', *pa-H<ar>uTHuT* 'make a crunching noise', *pa-k<ar>ungkung* 'make a noise of beating a box', *ng<ar>awngaw-an* 'make a noise (of people)' (Tsuchida 1983).

Example of *-ag-*: *T<ag>agTag* 'pour out' (Teng 2008:37).

Examples of *-a-*: *ki<a>pkip* 'eyelashes', *ku<a>skus* 'collarbone' (Teng 2008:37)*; hu<a>thut* 'front teeth', cf. *huthut* 'nibble', *ku<a>skus* 'shoulder blade', *tu<a>ktuk* 'hammer', cf. *tuktuk* 'hit with a hammer', *tu<a>HtuH* 'chisel', cf. *tuHtuH* 'to chisel', *da<a>ngdang* 'sickle', cf. *dangdang* 'cut with a sickle', *Ta<a>rTar-an* 'clapper, noise maker to scare birds away', *pa-Tu<a>ngTung-an* 'drum', cf. *pa-TungTung* 'play with a drum', *tu<a>aptup* 'beak', cf. *tuptup* 'to peck' (Tsuchida 1983).

Notice that the infix *-a-* is inserted between the first vowel and the following consonant in Puyuma. So it is a different type of infix.

Puyuma /l/ is historically derived from PAn *N or *n', /L/ from *l, /r/ from *R or *r, and /g/ from *g.

2.9 Tsou

Most Tsou dialects have lost PAn *l, *r, and *R, and dropped many vowels and consonants in antepenultimate syllables. It is, therefore, difficult to identify any of the nonproductive infixes in Tung's (1964) Tsou glossary. Fortunately /r/ was still present in the Duhtu dialect when Tsuchida investigated it in 1969. He has found only one fossilised form which contains an infix /r/ in the Duhtu dialect of Tsou, reflecting *-al-, *-aR- or *-ar-, as in: *p<r>e'pi'i* 'eyebrow' < **p<ar>i'pi'i* (pre-Tsou) < **pikpik* in pre-Tsouic.

2.10 Kanakanavu and Saaroa

2.10.1 Kanakanavu

There are three uncommon infixes in Kanakanvu, *-ar-, -al-, -an-*, as in:[15]

Examples of *-ar-*: *k<ar>askas-a* 'fish basket', *m-ari-p<ar>aipai* 'blow (of wind)', *m-uru-s<ar>angsang* 'trickle down, as grease from meat when grilled,' *um-a-ar-ivi* 'hold a baby in one's arm' < *aR-iba,[16] *v<ar>anvan-a* 'rainbow'.

Examples of *-al-*: *v<al>ang(i)vangi* 'fish sp.', *pa-t<al>u(ng)-tungku-an* 'kapok, plant sp., Bombax Ceiba, Linn'.

---

[15] Stress, which is phonemic in Kanakanavu, is left out in these examples.

[16] Compare Pazih *m-iba*, Bunun *m-al-iba, al-iba-an*, Atayal *g-m-iba* 'hold a baby in one's arm.'

Examples of *-an-*: *s<an>apisapi* 'driftwood', *ma-c<an>angcang* 'have a healthy complexion', *s<an>avikavika* 'fontanel', *k<an>eet-a* 'shortest bone in upper ribs' (cf. Maga Rukai *k<l>ɨtkɨt-a* 'clavicles'), *v<an>kavuka* 'trousers' (cf. Saaroa *v<an>ukavuka*).

Kanakanavu /r/ is derived from PAn *R, /l/ from *l or *y (in medial position preceded and followed by /a/), and /n/ from *n or *N.

2.10.2 Saaroa

There are four nonproductive infixes, *-al, -ar-, -alh-* and *-an-,* in Saaroa, as in:

Examples of *-al-*: *v<al>u'avu'a* 'large intestines', *c<al>ukacuka* 'cypress tree', *k<al>asekase* 'bamboo basket,' *k<al>ungkung-an* 'very skinny', *ma-s<al>angesange* 'beautiful', *ma-s<al>iisii* 'smooth', *pa-t<al>un-tungku-an* 'kapok, plant sp., Bombax Ceiba, Linn'.

Examples of *-ar-*: *v<ar>alhevalh-a* 'rainbow', *ma-k<ar>imkimi* 'salty', *lhi-ma-v<ar>asevase* 'hard-cooked rice, leftover rice', *um-a-ar-iva* 'hold a baby in one's arm'.

Examples of *-alh-*: *ma-t<alh>eketeke* 'cool', *ma-v<alh>auvau* 'healthy', *k<alh>eet-a* 'shortest bone in upper ribs' (cf. Maga Rukai *k<l>ɨtkɨt-a* 'clavicles').

Example of *-an-*: *v<an>ukavuka* 'trousers' (only one example).

Saaroa /l/ is historically derived from *l, /r/ from *R or *r, /lh/ from *N or *j, and /n/ from *n.

2.11 Atayal

There is a general vowel-deletion rule before stress in most Atayal dialects and the Atayalic language Seediq. So it is difficult to discern if there is an infix of *-al-*, *-ar-*, *-an*, etc. in their lexical forms without the vowels right after the initials. In the most conservative dialect of Mayrinax, there is a little evidence for the existence of these infixes when compared with the other Atayal dialects, as in:

Example of *-a-*: *w<a>ylung* 'chicken' in Mayrinax, cf. *wilung* in Sakuxan dialect of Atayal.

Examples of *-ag- ~ -g-*: *k<ag>isi'* 'basket carried on one's back' in Mayrinax, cf. *kisi'* in Sakuxan, *q<ag>ungu'* 'loom' in Mayrinax, cf. *qungu'* in Squliq dialect, *s<ag>asap* 'eaves' in Matabalay, cf. *sabsab*[17] in Mayrinax, *b<g>alaw* (< *b<aR>aNaR) 'Smilax China' in Pianan and Lmuan dialects, cf. *balag* (< *baNaR[18]) in Mayrinax.

Examples of *-al-*: *k<al>ahang-an* 'care for', cf. *khang-an* in Squliq (Egerod 1978:315).

Mayrinax /g/ is derived from PAn *R, /l/ from *N, and *l is lost. So *-a-* may reflect *-al-.

---

[17] Note the loss of medial *-b-* before another consonant in Matabalay and that /-b, -g/ in Mayrinax are devoiced in most other Atayal dialects.

[18] Compare *valja* in Paiwan, *banal* in Bunun, *blaa* in Maga Rukai, *vanare* in Kanakanavu, *valhare* 'Smilax China' in Saaroa, *banal* in Bontok (Madulid 2001), *banag* 'Smilax bracteata, Pres.' in Ilokano, and *banar* 'Smilax bracteata' in Malay.

However, the Atayalic group of languages also has the rare infixes, *-a-, -na, -in-, -il-,* and *-i-* in the male forms of speech in the Mayrinax dialect,[19] and they are inserted in the second syllable instead of the first syllable of the root, as in Li (1982, 1983):

Examples of *-a-* (before final consonant of stem): *q<um>alu<a>p/q<um>alup* 'hunt' < *qaNup, *'imu-a-g/'imug* 'house', *q<um>asu<a>g /q<um>asug* 'distribute', *h<um>aku<a>t/h<um>akut* 'carry' < *SakuC, *r<um>uru<a>g/r<um>urug* 'push', *quma-quma<a>h/quma-qumah* 'dry field' < *qumaH, *q<in>iri<a>ng/q<in>iring* 'wall,' *qau<a>g* 'bamboo' < *qauR.

Examples of *-na-* (before final consonant of stem): *s<um>ayu<na>g/s<um>ayug* 'substitute', *raqi<na>s* 'face' < *DaqiS, *ta-thaw<na>k-an/ta-thawk-an* 'seat'.

Examples of *-in-* (before the vowel of the final syllable of stem): *qas<in>ug/qasug* 'beast', *bus<in>uk* 'drunk' < *buSuk, *buh<in>ug* 'bow' < *busuR, *ma-ba<in>ay /ma-baay* 'buy' < *beli.

Examples of *-il-* (before the vowel of the final syllable of stem): *guq<il>uh/guquh* 'banana', *mat<il>uq/mataq* 'raw', *huk<il>u'/hauku'* 'plant sp.'[20]

Examples of *-i-* (before the vowel of the final syllable of stem): *luh<i>ung/luhung* 'mortar' < *Nesung, *rul<i>ug/rulug* 'top of tree', *tak<i>is/takis* 'knife'.

2.12 Rukai

Unless indicated otherwise, most of the following examples are based on the Maga dialect of Rukai:

Examples of *-al- ~ -l-*: *k<al>oko-a* 'fingernail' in Labuan dialect < *kuS_2kuS_2 (Dahl 1981:35), *g<al>awgaw* 'finger, toe' in Tanan, Labuan and Budai dialects of Rukai, *g<l>ogav-a, ma-b<l>ɨsbɨsɨ* 'cool', cf. *bsɨbsɨ* 'wind,' *b<l>ikbik-a* 'the soft part right below the ribs', *k<l>uskus-a* 'shinbone', *k<l>ɨtkɨt-a* 'clavicles' (cf. *k<an>eet-a* in Kanakanavu and *k<alh>eet-a* 'the shortest bone in the upper part of the ribs' in Saaroa).

Examples of *-ar- ~ -r-*: *b<r>ingbingi* 'edge', *ma-v<r>ingvingi* 'doubt, suspect' *ma-c<r>ukcuku* 'headache', *k<r>angkangɨ* 'cinnamon', *k<r>ongkong-a* 'type of omen bird', *ma-prɨkpɨkɨ* 'astringent', cf. *ma-p<aL>ekepeke* 'astringent' in Mantauran, *ma-p<r>osposo* 'improperly cooked rice due to the lack of water', *ma-s<r>apsapɨ* 'shallow', *ma-s<r>ɨnsɨnɨ* 'painful', *ma-s<r>ɨgsɨgɨ* 'get numb', cf. *ma-maa-s<aL>ege-sege* in Kocapongan, *te-s<r>esee* 'soil mixed with pebbles', *t<r>engtenge* 'wheel'.

---

19 We can discern these infixes by comparing the innovative male forms with their corresponding more archaic female forms or reconstructed forms. Some of the female forms are given right next to their corresponding male forms separated by a slash. In addition to the insertion of an infix in these male forms, there may be some other minor modifications in vowels or consonants.

20 Note the irregular vowel /u/ instead of /a/ in the male form *matiluq* and /u/ instead of /au/ in the male form *hukilu'*.

Let it be noted that Maga has lost the distinction between the flap /L/ and trill /r/ distinguished in most other Rukai dialects and has merged them as a single /r/. The flap /L/ reflects PAn *l while the trill /r/ reflects PAn *R (Li 1977). In other words, *-ar- ~ -r-* in Maga have two different sources, PAn *-al- and *-aR-, as attested in the other Rukai dialects, such as Mantauran and Kocapongan, as illustrated above.

Rukai /l/ is historically derived from PAn *N or *n', /L/ from *l, and /r/ from *R.

2.13 Tagalog

There is also evidence for these nonproductive infixes in extra-Formosan, such as *-al-*, *-ag-*, *-an-*, and *-a-* in Tagalog, as in (De Guzman 1966 and Institute of National Language 1940):

Examples of *-al-*: *h<al>akhak* 'laughter', *h<al>akhak-an* 'laugh at', *h<al>aghag* 'full of neglect', *h<al>imhim-an* 'hatch', *h<al>inghing* 'moaning', *h<al>ughug-in* 'ransack', *b<al>imbing* 'tree which produces acid edible fruit', *b<al>akubak* 'dandruff', *b<al>asubas* 'one who procrastinates', *b<al>agtas* 'cross a road' < *bagtas* 'cross a road', *b<al>anga'* 'earthen jar' < *banga'* 'native earthen jar', *k<al>adkad* 'drag' < *kadkad* 'unfold', *k<al>awkaw* 'stir water with the hand' < *kawkaw* 'dip the hand in the water', *k<al>ibkib* 'copra' < *kibkib* 'gnaw away coconut meat in the shell', *d<al>apdap* 'strip off', *g<al>awgaw* 'a restless person', *ng<al>itngit* 'creaking', *p<al>atpat* 'long, split bamboo' < *patpat* 'piece of split bamboo', *s<al>ansan* 'stack, heap, pile' < *sansan* 'repeatedly', *s<al>ingsing* 'ring of curtains' < *singsing* 'ring', *t<al>uktok* 'summit, top' < *tuktok* 'summit of a mountain'.

Examples of *-ag-*: *l<ag>ablab* 'blaze', *l<ag>aslas* 'noise made by a brook', *d<ag>angdang* 'expose to heat or fire' < *DangDang, *d<ag>ubdob* 'blaze', < *dubdob* 'build fire,' *d<ag>uldol* 'grind, strike' < *duldol* 'shove, thrust with force', *h<ag>alhal* 'outburst of laughter', *h<ag>ayhay* 'dry in the sun', *l<ag>ablab* 'blaze', *l<ag>aylay* 'hanging, drooping' < *laylay* 'hanging, drooping', *p<ag>akpak* 'flapping of wings' < *pakpak* 'wing' < *pakpak 'beat the wings', *p<ag>aspas* 'shaking' < *paspas* 'dust off, shake off the dust', *p<ag>aypay* 'flapping of wings, shaking' < *paypay* 'fan', *s<ag>alsal* 'continuous and strong spurting of liquid' < *salsal* 'masturbation', *s<ag>ansan* 'continuous' < *sansan* 'repeatedly', *s<ag>imsim* 'premonition' < *simsim* 'taste a little to prove the flavour', *s<ag>insin* 'close together' < *sinsin* 'closely woven', *s<ag>itsit* 'hissing sound', *s<ag>utsot* 'suck noisily' < *sutsot* 'whistle', *t<ag>aktak* 'downpour', *t<ag>inting* 'tinkling, jingling', *t<ag>istis* 'rapid dripping or leaking of liquid', *t<ag>uktok* 'brief, solid sound' < *tuktok* 'knock at the door', *w<ag>ayway* 'waving, fluttering of something' < *wayway* 'long piece'.

Examples of-*an-*: *s<an>aysay* 'essay' < *saysay* 'narration', *k<an>away* 'a marine bird entirely of white colour', cf. *kaway* 'call by waving the hand or handkerchief'.

Examples of *-a-*: *d<a>igdig* 'world', *t<a>imtim* 'devoted', *b<a>iki'* 'mumps' < *biki'* 'mumps'.

There are many forms that contain the infixes *-al-* and *-ag-*, which are not listed above, but only the two forms with *-an-*. Notice that the infix *-a-* also occurs in a few forms in Tagalog. However, this infix *-a-* is actually *-a'-* ,[21] which reflects *-al-, *-aq- or *-'-, e.g., *Zalan > *da'an* 'road', *CalingaH > *ta'inga* 'ear', *paqiC > *pa'it* 'bitterness', *ka'en > *k<um>a'in* 'eat' in Tagalog.

Tagalog /l/ is historically derived from *l, *r or *j, /g/ from *R or *g, and /n/ from *n or *N. Since reflexes of *-al-, *-aR- and *-aN- are common in Formosan languages, while reflexes of *-aj-, *-ar-, *-ag- or *-an are only rarely attested, if at all, we can infer that these Tagalog infixes reflect *-al-, *-aR- and *-aN- respectively.

2.14 Javanese

There are four nonproductive infixes *-al-*, *-el-*, *-er-* and *-r-*, in Javanese, as in Lopez (1977), Robson and Wibisono (2002):

Examples of *-al- ~ -l-*: *k<al>amangga ~ kemangga* 'spider', *l<al>awa ~ lawa* 'a certain small bat', *l<al>ilah* 'my God!' cf. *lilah* 'permission', *p<al>ikrama* 'honour, respect', cf. *pikrama* 'marry' (Robson and Wibisono 2002), *b<l>iya* 'make an opening' < *biyak, *g<l>emgem* 'silent in mourning', *t<al>utuh/t<l>utuh* 'flow on' (Lopez 1977).

Examples of *-el-*: *t<el>iti* 'parentage' < *taytay, *p<el>engkung* 'semi-circle' < *be(ng)kung, *k<el>entung* 'bird rattle' < *kentung, *b<el>enging* 'very clean premises' (Lopez 1977).

Examples of *-r-*: *t<r>ata* 'rattle, crattle' < *taktak, *t<r>etek* 'quake, quiver', cf. *tetek* 'knock, tap' < *tektek, *t<r>iti* 'pulverize', cf. *titi* 'tap' <*tiktik, *b<r>abah* 'speak loud' < *babah, *k<r>ekes* 'begged off' < *kiskis, *k<r>udug/k<l>udug* 'dull sound, thud' < *kuDug, *t<r>intjing* 'nimble' < *tingting (Lopez 1977).

Examples of *-er-*: *gesah/g<er>esah* 'sigh' < *kesaq, *tj<er>emed/tj<r>emed* 'unchaste, impure' cf. *tjemer* 'not clean' < *cemed, *t<er>antjang/t<r>antjang* 'transparent' (Lopez 1977).

Javanese /l/ is historically derived from PAn *l and /r/ from *r or *j. It can be assumed that both *-al- ~ -l-* and *-el-* reflect *-al-, while *-er-* and *-r-* reflect *-ar- or *-aj- because, as pointed out by an anonymous reviewer, Javanese /e/ is the usual reflex of PAn *a in antepenultimate position. That is to say, *-el-* and *-er-* are a later development in Javanese. Moreover, there is little Formosan evidence for the 'pepet original' in these infixes.[22]

---

[21] There is a glottal stop between two vowels in Tagalog.

[22] Naomi Tsukida (pers. comm.) provides us with a few lexical forms contaning *-el-, -er-* and *-ed-* in the Truku dialect of Seediq: *huyuq ~ h<el>uyuq* 'sharp', *hemadan ~ h<er>emadan* 'siblings of opposite sex', *kedekat ~ ked<er>ekat* 'scales of fish', *be'nux ~ b<er>e'nux* 'plains', *selut ~ s<ed>elut* 'attached'. These infixes appear only in antepenultimate syllables, and the vowels have been reduced. Although we do not know what the original vowels were, we may assume that, similar to Javanese, Seediq *e* is the usual reflex of PAn *a in these forms. If so, then Seediq *-el-, -er-* and *-ed-* reflect PAn *-aN-, *-al- and *-aD- respectively.

3 Summary and discussion

The nonproductive infixes of the form of *-aC-* (or its variants *-a-* or *-C-*) occur in all Formosan languages and many western Austronesian languages belonging to all the primary subgroups of Austronesian, following the classification by Blust (1999) and Ross (2009). Lexical forms containing these infixes may be numerous in some languages, such as Paiwan, Amis, and Tagalog, or very few in other languages, such as Tsou, Saisiyat, Basay, and Pazih. This discrepancy may be partly due to the lack of detailed dictionaries for the latter, such as Tsou and Saisiyat. It is not surprising that we have found very few examples for Favorlang, Siraya, and Basay, the first two extinct languages with limited written documents from the 17th century, and the last one with limited field notes by Asai in 1937 (Asai 1991).

Reflexes of the nonproductive PAn affixes in Formosan languages and two western Austronesian languages are as shown in Table 1.

Table 1: Reflexes of three PAn infixes

PAn	*-al-	*-aR-	*-aN-
Thao	*-ar-*	–	*-az-*
Favorlang	*-ar-*	–	–
Kavalan	*-ar-,-aR-*	*-al-,-ar-, -aR-*	*-an-*
Basay	–	*-al-,-ar-*	*-an-*
Siraya	*-ar-*	–	*-al-*
Amis	*-al-*	*-al-*	*-alh-*
Saisiyat	*-aL-*	*-aL-*	*-al-*
Pazih	*-ar-*	–	*-al-*[23]
Bunun	*-a-*	*-al-*	*-an-*
Paiwan	*-al-*	*-a-*	*-alj-*
Puyuma	*-aL-*	*-ar-*	*-al-*
Tsou	*-r-*	*-r-*	–
Kanakanavu	*-al-*	*-ar-*	*-an-*
Saaroa	*-al-*	*-ar-*	*-alh-*
Atayal	*-a-*	*-ag-*	*-al-*
Rukai	*-aL-*	*-ar-*	*-al-*
Tagalog	*-al-, -a'-*	*-ag-*	*-an-*
Javanese	*-al-, -el-*	–	–

Reflexes for *-al-, *-aR- and *-aN- are quite common in Formosan languages as well as western Austronesian languages although we have found only two western Austronesian languages, Tagalog (see §2.13) and Makassarese (south Sulawesi), with reflexes for *-aN-.[24] Reflexes for *-aD-, *-ar-, *-ag-, *-aq-, and *-ay- are attested only in one Formosan language each. Reflexes for *-an- are unambiguously attested in only one Formosan language, Amis,

[23] There is only one dubious example of Pazih *-al-* reflecting *-aN-: *k<al>ikux* 'fingernail' < *kuS_2kuS_2; see §2.5.2.

[24] Makassarese (Cense 1979) examples of *-an-*: *k<an>eke* 'edible seebass' (cf. *keke* 'small'), *k<an>ingking* 'little finger or toe', *k<an>uku* 'nail' (cf. *k<alj>uskus-an* in Paiwan, *ka-k<l>okoeh* in Saisiyat).

with only a few examples; they are ambiguous in Kavalan, Bunun, Kanakanavu and Tagalog because /n/ is historically derived from either *n or *N in these languages. Reflexes for the infix *-al- are attested in all Formosan languages except Basay, including *-ar-* in Thao, Kavalan, Rukai (Maga), Pazih, Favorlang and Siraya, *-al-* in Amis, Paiwan, Kanakanavu and Saaroa, *-aL-* in Rukai (other than Maga), Saisiyat and Puyuma,[25] *-a-* in Bunun and Atayal (Mayrinax), and *-r-* in Tsou (Duhtu). Reflexes for *-aR- and *-aN- are also found in a great number of Formosan languages as well as Tagalog,[26] as shown in Table 1. We may conclude that these three infixes, *-al-, *-aR-, and *-aN-, can be reconstructed for Proto Austronesian, but not the other five infixes mentioned above. These three infixes, nonproductive in contemporary languages, may have been productive in Proto Austronesian, but fallen out of use for one reason or another.

Dempwolff's 'erstarrte Infixe' (petrified infixes) include *al-, *-ar- and *-aR-. After listing many forms containing nonproductive infixes in mostly western Austronesian languages, Lopez (1977) concludes that 'evidences [sic] point to the justification and validity of retaining Dempwolff's construction of *-aR-, *-al- and *-ar-, to which may now be added *-el and *-er-.' However, Lopez's study is mainly based on western Austronesian languages, and there is little or no Formosan evidence for reconstructing *-el- or *-er-. Blust (2009:380) also concludes that 'there appears to be comparative evidence for three phonologically similar but distinct infixes *-al-, *-ar- and *-aR-'. However, his evidence for reconstructing *-ar- is based mostly on western Austronesian languages, which all belong to the same major subgroup, Malayo-Polynesian, and he cites very little evidence from the other subgroups: only two examples of *-ar-* in Pazih and also two in Paiwan[27]. In fact, *-ar-* in Pazih most generally reflects *-al-, rather than *-ar- (see Blust 1999b:333–334). More solid evidence is necessary to reconstruct *-ar- for PAn.

The main problem of reconstructing one or more infixes in the shape of *-aC- for Proto Austronesian is that it is difficult to reconstruct lexical items attested in two or more languages distantly related to each other. Nevertheless, we do find a few fossilised forms in distantly related Formosan languages, such as *t<an>uduq* 'finger' in Bunun, *ts<alj>udjuq* 'finger' in Paiwan, and *t<ar>olho'* < *t<alh>oro'* 'finger' in Amis < *tuZuq, all reflecting the same infix, *-aN-. Moreover, Lopez (1977) identified many lexical forms with these nonproductive infixes in many western Austronesian languages with PAn reconstructed forms.

Another problem is to determine the function of each of these infixes. The function of *-aR- is probably 'distributive, plural,' as suggested by Reid (1994). The function of *-aN- seems to be 'having the sound or quality of,' as reflected in Paiwan and Puyuma, or that of *-al-, as reflected in Amis and Tagalog. In fact, many other languages also have infixed forms with a similar function, but reflecting some other infixes. Still another function is 'instrument,' such as *-a-* in Puyuma, which is inserted after the first vowel, but it represents a different type of infix.

---

[25] In fact, *-al-* in Amis, Paiwan, Kanakanavu and Saaroa, *-aL-* in Rukai, Saisiyat and Puyuma are phonetically similar: Both *l* and *L* represent the flap, and the difference is only one of orthography.

[26] Reid (1994:330) also states that 'reflexes of … *-aR- 'distributive, plural' are found in many [Austronesian] languages,' but no example is given.

[27] In fact, there are many examples of *-ar-* in Paiwan (see §2.7), which might be derived from *-ar-, but unfortunately there are few lexemes which are clearly derived from *r. We, therefore, cannot be certain about this problem.

These nonproductive infixes may have been productive in PAn, but lost their productivity and functions at an early stage. Also they may have already been fossilised in PAn lexemes, just as the productive infix *-in-, which can be reconstructed as a non-productive element with some lexemes, e.g., *C<in>aqi 'guts' < *Caqi 'excrement.' Compare *-ar- 'distributed action' in *k<ar>uSkuS 'to scrape' < *kuSkuS 'finger', as proposed by Sagart (1994).[28] Most reconstructed PAn stems are dissyllabic, while only a few are trisyllabic, such as *Calinga 'ear', *qaRisam 'stem of cogon grass,' and *qaNingu 'shadow, reflection.' These forms may contain the infixes *-al-, *-aR-, and *-aN-[29] respectively. Similar forms are *baluku 'winnowing basket', *balija 'batten, reed of a loom', *da[r]izi 'finger', *baRius 'storm wind', *zaRami 'rice straw',[30] *qaNiCu 'ghost', and *paNiki 'bat'.[31]

It is clear that these nonproductive infixes require further study.

References

Adelaar, K. Alexander. 2000. Siraya reduplication. *Oceanic Linguistics* 39(1).33–52.

Asai, Erin. (ed. by Tsunekazu Moriguchi). 1991. Basai vocabulary. In Shigeru Tsuchida, Yukihiro Yamada and Tsunekazu Moriguchi, eds *Linguistic materials of the Formosan Sinicized populations I: Siraya and Basay*, 195-229. Tokyo: University of Tokyo.

Blust, Robert. 1999a. 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*, 3194. Symposium Series of the Institute of Linguistics (Preparatory Office), No. 1. Taipei: Academia Sinica.

——— 1999b. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2).321–368.

——— 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.

——— 2003. *Thao dictionary*. Institute of Linguistics, Taipei: Academia Sinica.

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

---

[28] Sagart (1994:275–277) reconstructs PAn *-ar-, as based on the infix *-ar-* found in Paiwan and Amis, *-al-* in Tagalog, and *-er-* in Malay. While Paiwan /r/ reflects PAn *r, Amis /r/ reflects PAn *D and Tagalog /l/ reflects PAn *l. Paiwan is the only Formosan language that reflects PAn *-ar-. Amis has many lexical forms containing the infix *-al-* (not cited by Sagart), which reflects PAn *-al- or *-aR-; see §2.4.

[29] It is suggested that there might be a PAn prefix *qaNi- 'spiritual being' in the forms such as *qaNingu 'image' and *qaNiCu 'ghost,' as well as *qali/kali- (Blust 2001). Blust (pers. comm.) does not believe from his study of the *qali/kali- prefixes in Austronesian languages comprise these infixes.

[30] Compare *Rami* in Kavalan, which may have lost the first syllable and *djamia* in Paiwan in which there may be a metathesis.

[31] Some other trisyllabic reconstructed forms contain an empty vowel between two syllables, e.g., *e in *CuqelaN 'bone', *baqeRuH 'new', *qaSeluH 'pestle', *suRekuD 'stick', *qiCeluR 'egg', *bineSiq 'seed', *ma-NiSepis 'thin', *ma-kaSepal 'thick', *qalejaw 'day,' *angetaq 'raw', *u in *panguDaN 'pandanus', or *a in *dakadak 'kick'. A few other forms contain an optional initial syllable *qa, as in *(qa)lipan 'centipede', *(qa)Nuang 'deer', or a suffix *-an, such as *qamiS-an 'north wind'.

Campbell, Rev. William. 1896. *The articles of Christian instruction in Favorlang Formosan, Dutch and English from Vertrecht's manuscript of 1650*. London: Kegan Paul, Trench, Trubner and Co.

Cauquelin, Josiane. 2008. *Ritual texts of the last traditional practitioners of Nanwang Puyuma*. Institute of Linguistics, Taipei: Academia Sinica.

Cense, A.A. 1979. *Makassaars-Nederlands Woordenboek*. 's-Gravenhage: Martinus Nijhoff.

Dahl, Otto Christian. 1981. *Early phonetic and phonemic changes in Austronesian*. Oslo: Instituttet for Sammenlignende Kulturforskning.

De Guzman, Maria Odulio. 1966. *English–Tagalog and Tagalog–English*. Manila: National Book Store, Inc.

Dempwolff, Otto. 1934–38. *Vergleichende Lautlehre des Austronesischen Wortschatzes*. Berlin: Dietrich Reimer.

Egerod, Søren. 1978. *Atayal–English dictionary*. London: Curzon Press.

Ferrell, Raleigh. 1982. *Paiwan dictionary*. Canberra: Pacific Linguistics.

Fey, Virginia. 1986. *Amis dictionary*. Taipei: The Bible Society.

Institute of National Language. 1940. *Tagalog–English vocabulary*. Manila: Institute of National Language.

Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology (BIHP)* 48(1).1–92.

—— 1978. A comparative vocabulary of Saisiyat dialects. *BIHP* 49(2).133–199.

—— 1982. Male and female forms of speech in the Atayalic group. *BIHP* 53(2).265–304.

—— 1983. Types of lexical derivation of men's speech in Mayrinax. *BIHP* 54(3).1–18.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Taipei: Academia Sinica.

—— 2006. *Kavalan dictionary*. Taipei: Academia Sinica.

Lopez, Cecilio. 1977. Nonproductive infixes in Indonesian. In Ernesto Constantino, ed. *Selected writings of Cecilio Lopez in Philippine linguistics*, 211–226. Quezon City: Archives of Philippine Languages and Dialects, Philippine Linguistics Circle.

Madulid, Domingo A. 2001. *A dictionary of Philippine plant names*. Manila: Bookmark, Inc.

Nihira, Yoshiro. 1988. A Bunun vocabulary, a language of Formosa. Unpublished manuscript. Tokyo.

Reid, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33(2).323–344.

Robson, Stuart and Singgih Wibisono. 2002. *Javanese-English dictionary*. Hong Kong: Periplus Editions Ltd.

Ross, Malcolm D. 1995. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Cheng-Hwa Tsang,Ying-Kuei Huang, Dah-An Ho and Chiu-Yu Tseng, eds *Austronesian studies relating to Taiwan*, 727–791. Taipei: Academia Sinica.

—— 2009. The primary subgroups of Austronesian: a reappraisal. In Alexander Adelaar and Andrew Pawley, eds *Austronesian historical linguistics and culture history: a festschrift for Robert A. Blust,* 295–326. Canberra: Pacific Linguistics.

Sagart, Laurent. 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics* 33(2).271–308.

Taintor, E.C. 1874. The aborigines of northern Formosa. *Journal of the North China Branch of the Royal Asiatic Society* N.S. 9:53–88.

Teng, Stacy F. 2008. *A reference grammar of Puyuma: an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.

Tsuchida, Shigeru. 1962. Saisiyat vocabulary. Unpublished manuscript. Tokyo: Tokyo University.

—— 1980. Puyuma (Tamalakaw dialect) vocabulary, with grammatical notes and texts, *The peoples, cultures and languages of the Black Currents*, 183–307. Tokyo: Kadokawa Shoten.

—— 1983. Puyuma-English index (Austronesian language in Taiwan). *Working Papers in Linguistics '83*. 83.10–63. Tokyo: University of Tokyo.'

Tung, T'ung-ho. 1964. *A descriptive study of the Tsou language, Formosa*. Taipei: Academia Sinica. (Reprinted in 1991 with Notes and Errata prepared by Paul Jen-kuei Li.)

Utrecht manuscript. 1600s. Vocabularium Formosanum. Utrecht University Library. (Reproduced in Naojiro Murakami. 1933, *Sinkang manuscripts*, 154–202. Taihoku, Formosa: Taihoku Imperial University.)

Wolff, John. 1973. Verbal inflection in proto-Austronesian. In Andrew B. Gonzales, ed. *Parangal kay Cecilio Lopez,* 71–91. Manila: Linguistic Society of the Philippines.

Zeng, Si-qi. 2008. *A handbook of productive affixes and verb roots in Amis*. Beijing: Minority Peoples Publisher. [in Chinese]

19 *Proprietives in Oceanic*

FRANTISEK LICHTENBERK

1 Introduction[1]

In 2000 Malcolm Ross published a paper (originally delivered in 1999 at the Fourth International Conference on Oceanic Linguistics) that dealt with, among other things, the Proto Oceanic (POc) suffix *-[k]a (Ross 2000). He suggested there that the function of the suffix was 'to derive adjectival nouns from both nouns and verbs' (2000:332).

The present study deals with the same morpheme but from a somewhat different perspective, and some of the conclusions reached here differ from those reached by Ross. I will return to his study in §5.

In grammars of many Oceanic languages one finds references to a suffix typically added to nouns that serves to derive stative verbs or adjectives. The semantic contribution of the suffix is characterised in various ways, but typically, although not always, the derived verb or adjective is said to signal that an entity, place, etc. has or contains a lot of whatever the source noun designates; cf. English *dust* and *dusty (furniture)*. I will refer to the morpheme with this function as 'proprietive' (PROP). Examples of derivations using a proprietive suffix are given in (1)–(3).

Salisbury (2002:131) characterises the proprietive suffix *-a* in Pukapukan thus: 'This suffix derives a stative verb from a noun meaning "full of N", or "infested with N"'.

PUKAPUKAN

(1) *namu-a*[2]
mosquito-PROP
'mosquito ridden' (Salisbury 2002:131)

For the cognate suffix *-a* in Samoan, Mosel and Hovdhaugen (1992) use the term 'ornative'. It derives non-ergative verbs from nouns.

[1] It gives me great pleasure to offer this paper to Malcolm Ross. I am grateful to Malcolm for providing me with information on Takia, not knowing what the information was going to be used for. I also wish to thank two anonymous referees and the editor, Bethwyn Evans, for valuable comments on an earlier version of this paper.

[2] The Leipzig glossing conventions are used here, with the addition of the following abbreviations: NFUT non-future; PART partitive; PERS personal (suffix); PROP proprietive; RDP reduplication.

SAMOAN
(2) *vai-a*
water-PROP
'watery' (Mosel and Hovdhaugen 1992:205)

Tamambo has the suffix *-ha*, '[which] derives intransitive verbs which are "NOUN-like"' (Jauncey 1997:139).

TAMAMBO
(3) *banoi-ha*
volcanic.ash-PROP
'be covered with ash' (Jauncey 1997:139)

Although Jauncey characterises the meaning of intransitive verbs formed by means of *-ha* as 'NOUN-like', in most of the examples she gives the meaning as 'have or contain that designated by the base noun'. Nevertheless, as will be shown in later sections, meanings of the type 'NOUN-like' are found in many Oceanic languages, including Tamambo. It will be shown that there are two types of meaning associated with proprietive forms: (i) 'have, contain (an abundance of) X', 'having, containing (an abundance of) X'; and (ii) '(be) X-like', 'have/having properties, characteristics of X, without containing X', X being the kind of entity, matter, etc. designated by the base. For convenience, the two types of meaning will be represented as 'have, contain X' and 'be X-like', respectively.

The first three examples above illustrate the first type, and example (4) below from Toqabaqita the second type.

TOQABAQITA
(4) *uuka-qa*
Derris.sp-PROP
'taste very bitter (e.g. chloroquine)' (*uuka* designates a species of Derris climber that is pounded and used to stun fish in the sea) (Lichtenberk 2008a)

The purpose of the present study is investigation of proprietive forms and their functions in Oceanic languages and in POc. Cognate proprietive affixes are found in (at least) three different primary subgroups of Oceanic: Central/Eastern Oceanic, Western Oceanic, and in the Mussau language.[3] These are discussed in §2, §3 and §4, respectively. Section 5 deals with the proprietive morpheme in POc and summarises the findings.

As will be shown in §5, the POc proprietive morpheme had two variant forms, *-ka and *-a. In this study the main focus is on proprietive derivations that employ a reflex or reflexes of the POc morpheme. Also included are proprietive suffixes that are in some kind of alternation with the reflexes of the POc forms. (As briefly mentioned at the beginning of §5, some languages have proprietive derivations that employ other means.)

For the data given in what follows, page references are given only when the data are not easily located in the sources. Thus, page references are not given for examples that come from dictionaries where the proprietive form is listed as a separate headword.

[3] I assume a fairly conservative subgrouping of Oceanic here; see Lynch et al. (2002:94). For a different view of the primary subgroups of Oceanic see Ross et al. (2008). Adopting the latter subgrouping would not affect the reconstruction of the proprietive morpheme; in fact, it would make it stronger.

2 Proprietives in Central/Eastern Oceanic

2.1 Southeast Solomonic

Proprietive forms are common in the Southeast Solomonic subgroup. The discussion begins with Toqabaqita, the language for which I have the most detailed information.

2.1.1 Toqabaqita

2.1.1.1 The forms

The discussion is based on §4.2.2.3 in Lichtenberk (2008b). Some data can also be found in Lichtenberk (2008a). Unless specified otherwise, the data cited here come from the latter.

The proprietive morpheme in Toqabaqita has two allomorphs, *-qa* and *-la*, the former being a reflex of POc *-ka. (Orthographic *q* represents the glottal stop.) The distribution of the two allomorphs is partly determined phonologically: *-qa* is not used with bases that contain *q* in their last syllable, and, few exceptions apart, *-la* is not used with bases that contain a liquid, *l* or *r*, in their final syllable. (For examples of exceptional forms see [11b] and [15b].) Elsewhere, the distribution of the allomorphs is not predictable. In most such cases only one variant is possible, but there are also a number of bases that can take either variant in free variation. There is (at least) one case where a base can take either allomorph with a secondary difference in meaning.

The next three examples illustrate the phonological conditioning: *-la* with a base that contains *q* in the final syllable in (5), and *-qa* with a base that contains *l* in the final syllable in (6) and with a base that contains *r* in the final syllable in (7).

TOQABAQITA

(5) *feefeqa-la*
eye.mucus-PROP
'of eye(s): contain mucus'
(Note: *Feefeqa* 'eye mucus' is a reduplication of *feqa* 'defecate'.)

(6) *ula-qa*
vein-PROP
'of a body part: show veins, be veiny'

(7) *kuburu-qa*
storm-PROP
'of the sea: be rough, stormy'

Both in (8) and in (9) below, the consonant in the final syllable of the base is *f*, but in the former the proprietive suffix has the form *-la*, and in the latter *-qa*. Furthermore, the bases in the two forms are related: *ifu* 'hair, feather, fuzz (of plants)' and *ififi* 'stringy fibre (e.g. of mango)', the latter being an irregular reduplication of the former.

TOQABAQITA

(8) *ifu-la*
hair-PROP
'of a person's body: be hairy, hirsute'

(9) *ififi-qa*
fibre-PROP
'contain (a lot of) stringy fibres; be stringy'

In some cases either variant of the proprietive suffix is possible, without any semantic difference, as in (10). *Maamako-qa* is more common than *maamako-la*.

TOQABAQITA

(10) a. *maamako-qa*
mud-PROP
'be muddy'

b. *maamako-la*
mud-PROP
'be muddy'

There are three words for 'sea', 'sea water', 'salt', *qamali*, *asi* and *kwaimoli* (the last one not common). *Qamali* can take either variant of the suffix. Exceptionally it can take the *-la* variant, even though it contains *l* in the final syllable. *Asi* can take only the *-la* variant, and *kwaimoli* only the *-qa* variant.

TOQABAQITA

(11) a. *qamali-qa*
sea/sea.water/salt-PROP
'be salty', 'contain salt or sea water'

b. *qamali-la*
sea/sea.water/salt-PROP
'be salty', 'contain salt or sea water'

c. *asi-la*
sea/sea.water/salt-PROP
'be salty', 'contain salt or sea water'

d. *kwaimoli-qa*
sea/sea.water/salt-PROP
'be salty', 'contain salt or sea water'

In some cases, the proprietive derivation involves reduplication of the source noun.

TOQABAQITA

(12) *foo-fote-la*
RDP-paddle-PROP
'of a person or an animal: be thin, skinny'

There are a few pairs of proprietive forms one of which contains a non-reduplicated base and the other one a related reduplicated base, the two derived verbs being different in meaning. The two so-related proprietive forms may, but need not, contain the same variant of the proprietive suffix. In (13) they do, while in (14) they do not.

TOQABAQITA

(13) a. *fau-la*
stone/rock-PROP
'of ground: be stony, rocky'

b. *(lio) fau-fau-la*
look RDP-stone/rock-PROP
'be serious, stern', 'be/look stone-faced'

(14) a. *qabu-la*
blood-PROP
'be bloody, covered with blood'

b. *qabu-qabu-qa*
RDP-blood-PROP
'be of a shade of red, the colour of blood'

See also *wela-qa* in (27) and *wela-wela-qa* in (34) in §2.1.1.2.

There is (at least) one pair of proprietive forms that have the same base, one of which takes the *-qa* variant of the proprietive suffix, and the other the *-la* variant, and there is a difference in meaning.

TOQABAQITA

(15) a. *kale-qa*
egg-PROP
'of a bird: lay eggs'
(i.e., 'have eggs')

b. *kale-la*
egg-PROP
'lay a large number of eggs; normally used only about turtles,
which lay large numbers of eggs'

Kale-la is another instance of a phonologically exceptional form: the base contains *l* in the final syllable.

Most commonly, the proprietive suffixes are added to nouns, as in all the examples above, including compounds.

TOQABAQITA

(16) *uli-qai-la*
branch-tree-PROP
'of a traditional priest: wear an *uliqai* headdress'
(Note: The *uliqai* headdress consisted of branchlets from a certain tree.)

In one case the base is a noun phrase outside the proprietive form.

TOQABAQITA

(17) *teqe-bali-qa*
one-side-PROP
'be asymmetrical, lopsided'

There are a few cases of proprietive verbs where the base itself is an intransitive verb, usually, although not always, reduplicated. Whether in such cases there used to be a related noun is not possible to tell.

TOQABAQITA

(18) *bii-biinga-la*
RDP-sleep-PROP
'be sleepy'

Bii-biinga-la ‘be sleepy’ takes as its subject a noun phrase that designates the person’s or animal’s eyes. On the other hand, *biinga* ‘sleep’ takes as its subject a noun phrase designating the person or the animal sleeping.

Another example of a proprietive form with a verbal base is given in (19).

TOQABAQITA

(19) *en-eno-qa*
be.quiet/calm-RDP-PROP
‘of a place or a person: be peaceful’

There are a few pairs of derivationally related verbs, one without and one with the proprietive suffix, without any apparent difference in meaning. For example:

TOQABAQITA

(20) a. *qakweo*
‘of fingers and toes: be numb, stiff, cramped’

b. *qakweo-qa*
-PROP
‘of fingers and toes: be numb, stiff, cramped’

In at least one case the form with the suffix is intensive in its meaning, compared to the suffixless form.

TOQABAQITA

(21) *saa-sadi-la*
RDP-be.hard-PROP
‘be very hard’

As will be seen later, in some languages proprietive forms have or may have an intensive meaning. However, in (21) there is reduplication of the base *sadi* ‘be hard’, and it is not clear whether the intensive force of *saa-sadi-la* ‘be very hard’ is due to the presence of the proprietive suffix or to the reduplication (or to both).

2.1.1.2 The semantics

There are only a few proprietive verbs based on intransitive verbs, and no semantic pattern is discernible. The discussion that follows will, thus, focus on proprietive verbs based on nouns. Here two main patterns can be distinguished. In one, the meaning of a proprietive verb is, approximately, ‘have, contain X’, ‘be characterised by/noteworthy because of X’, where X is the kind of entity designated by the base noun. A few examples are given below.

TOQABAQITA

(22) *Qae-ku* *e* *karu-qa.*
leg-1SG.PERS 3SG.NFUT scar-PROP
‘My leg(s) is/are scarred.’/‘I have (many) scars on my leg(s).’

(23) *Si* *kula* *e* *thato-la.*
PART place 3SG.NFUT sun-PROP
‘The place is sunny.’

(24) *Botho e mariko-qa.*
pig 3SG.NFUT flesh-PROP
'The pig has a lot of (lean) flesh.'/'The pig is well-fleshed.'

(25) *Kisina e thasu-la.*
kitchen/cookhouse 3SG.NFUT smoke-PROP
'The kitchen/cookhouse is smoky/full of smoke.'

In (22)–(25) the relevant entity or place is characterised as having, containing (a lot of) X. But this is not necessarily so. Compare (25) above and (26) below.

TOQABAQITA

(26) *Era e thasu-la.*
fire 3SG.NFUT smoke-PROP
'The fire is giving off a lot of smoke.'

Here the more general semantic description 'be characterised by/noteworthy because of X' is more appropriate.

More often than not, a proprietive verb expresses the fact that the relevant entity, place, etc. is characterised by a lot of, an abundance of, too much of, what the base noun designates, as in (22)–(26). This, however, is not necessarily the case, as in (27).

TOQABAQITA

(27) *Roo ai kera wela-qa naqa.*
two woman 3PL.NFUT child-PROP PRF
'The couple have a child/children now.' (They are not childless anymore.)
(Note: The basic senses of *ai* are 'woman', 'wife', but the expression *roo ai* designates a married couple.)

The sentence in (27) can be used even if the couple have only one child. They are characterised by, noteworthy because of, having one or more children, as opposed to being childless.

Similarly, the sentence in (28) could be used even if the Malay apple contained just one worm. What is noteworthy about the Malay apple is that it contains one or more worms.

TOQABAQITA

(28) *Fa kabirei e waa-la.*
CLF Malay.apple 3SG.NFUT worm-PROP
'The Malay apple has (a) worm(s) inside it.'/'The Malay apple is wormy.'

The other basic meaning of proprietive verbs in Toqabaqita is to express the fact that the relevant entity, place, etc. is in some crucial respect X-like, without having, containing X. A few examples are given below.

TOQABAQITA

(29) *Roketa e taatada-la.*
chilli.pepper 3SG.NFUT tree.sp-PROP
'Chilli peppers are hot (pungent).'

Taatada designates a tree, *Dendrocnide* sp., with leaves that sting.

TOQABAQITA

(30) *Fo ongi e angi keekee-qa.*
CLF panpipe 3SG.NFUT make.sound cicada-PROP
'The panpipe played alto (lit. cicada-like).'

See also *uuka-qa* 'taste very bitter' in (4) in §1 (*uuka* '*Derris* sp.), and *foo-fote-la* 'of a person or an animal: be thin, skinny' ('be canoe-paddle-like'; *fote* 'paddle') in (12) in §2.1.1.1.

Included here are also a number of colour terms, whose meanings can be characterised as 'have the colour of X'.

TOQABAQITA

(31) *kaakaa-qa*
(white)cockatoo-PROP
'be white, esp. pure white'

(32) *buu-bulu-qa* / *bulu-bulu-qa*
RDP-black.pig-PROP / RDP-black.pig-PROP
'be black; be dark-coloured: dark grey, dark blue, dark purple'

See also *qabu-qabu-qa* 'be of a shade of red, the colour of blood' in (14b) in §2.1.1.1.

There is (at least) one colour proprietive verb whose base is a verb.

TOQABAQITA

(33) *maa-marakwa-qa*
RDP-be.green-PROP
'be green'

While both *marakwa* and *maa-marakwa-qa* mean 'green', the former, unlike the latter, is used primarily in compounds.

There are also colour terms that historically most likely contain the proprietive suffix, but the presumed base does not exist in present-day Toqabaqita, such as *meemenaqa* 'be red/purple'.

As is typical of derivation in general, there are also proprietive verbs that do not fit neatly in either semantic category. For example:

TOQABAQITA

(34) *wela-wela-qa*
RDP-child-PROP
'of children and young banana trees (before fruiting):
have grown well; have a good, strong, well-developed body'

Compare *wela-qa*: 'of a married couple: have one or more children'; see (27) above.

TOQABAQITA

(35) *lalamoa-qa*
victim-PROP
'of a person's behaviour: be so bad that the person deserves to be killed'
('behaviour/behaving' as subject)

Lalamoa designates a person who died a violent death (e.g. in a fight, rather than of sickness or old age), a victim of murder, fight.

We can now move to proprietive forms elsewhere in Southeast Solomonic.

2.1.2 Proprietives in other Southeast Solomonic languages

Proprietive forms are found in many Southeast Solomonic languages and the survey that follows is by no means exhaustive. Lau has the suffixes *-a* and *-la*, which Fox (1974:1, 110) calls 'adjectival'.

LAU

(36) *rodo-a*
night-PROP
'dark' (Fox 1974:163)

(37) *'ae-la*
leg-PROP
'rooted, firm, sure, having legs'

The form *-la* is cognate with Toqabaqita *-la*, and *-a* is most likely cognate with Toqabaqita *-qa* (see note a in Table 1 in §5), but Fox does not further comment on the use of the two forms. There is at least one base that can take either suffix, with some difference in meaning. One meaning is of the 'have, contain X' type, and the other of the 'be X-like' type.

LAU

(38) *'abu-la*
blood-PROP
'bloody'

(39) *'abu-a*
blood-PROP
'red, crimson, Prussian blue (colour of deep sea)', 'bloody' (Fox 1974:3)

Fox also mentions two other suffixes that he calls adjectival, *-ra* and *-na*. In (40) the suffix *-ra* is added to bases that are reduplications of the verb 'fear'.

LAU

(40) *mou-mou-ra / mo-mou-ra*
RDP-fear-PROP / RDP-fear-PROP
'afraid', 'dizzy' (Fox 1974:138)

Besides *mou-mou-ra* and *mo-mou-ra*, Fox (1974:138) also gives *mou-mou-la* and *mo-mou-la* 'afraid'.

As an example of the form *-na* Fox (1974:139) gives *babalana*, without a gloss, either for the base or for the derived form.

Kwara'ae has the suffixes *-la* and *-'a*, used to form adjectives from verbal and nominal bases. Deck (1934) does not comment on the distribution of the two forms, but it is likely that phonological conditioning of the kind found in Toqabaqita (§2.1.1.1) applies. Deck gives three proprietive forms with *l* in the final syllable of the base and one with *r* in the final syllable of the base, and in all four cases it is the *-'a* variant that is used. In Kwara'ae too both types of proprietive meaning are found.

KWARA'AE

(41) *fau-la*
stone-PROP
'stony'

(42) *moli-moli-ʻa*
RDP-lemon-PROP
'spherical' (Deck 1934:34)

Deck (1934:34) gives one form that may contain the two proprietive suffixes simultaneously, and draws the reader's attention to it : 'note two suffixes'. However, the base is a verb, and it is possible that the suffix *-la* is the nominalizer rather than the proprietive suffix.

KWARAʻAE
(43) *oga-la-ʻa*
want-PROP?/NMLZ?-PROP
'selfish'

However, there is at least one Southeast Solomonic language, Ulawa, in which the proprietive morpheme can be present twice in one form; see examples (52) and (53) further below.

For Kwaio, Keesing (1985:76) gives the suffixes *-la* and *-ʻa*, 'which operate in parallel fashion, convert[ing] nouns to statives [i.e., stative verbs, F.L.]'.

KWAIO
(44) *fou-la*
stone-PROP
'be stony'

(45) *ʻola-ʻa*
thing-PROP
'be rich' (Keesing 1985:76)

For *ʻola-ʻa* 'be rich' compare Toqabaqita *donga-la* 'be rich; possess many things', based on *donga* 'bits and pieces of things in one place; all the parts, components of something', and Lau *donga* 'thing (name forgotten)', 'belongings' (Fox 1974).

In Kwaio the proprietive suffixes can be added to a few stative bases, in which case 'the resulting stative form may be substantially modified semantically, or little changed in meaning from the original base' (Keesing 1985:76):

KWAIO
(46) *gwari-ʻa*
be.cold-PROP
'be cured' (Keesing 1985:77)

(47) *tege-la*
be.strong-PROP
'be strong' (Keesing 1985:76)

Keesing does not comment on the conditions on the use of the two suffixes, but there may be some phonological conditioning, not unlike the conditions on the use of *-qa* and *-la* in Toqabaqita (§2.1.1.1). In all four examples given by Keesing that have *l* or *r* in the final syllable of the base it is the *-ʻa* variant that is used.

For ʻĀreʻāre, Geerts (1970) lists several forms that contain a proprietive suffix. For example:

ʻĀREʻĀRE

(48) *hau-ʻa*
stone/rock-PROP
'stony'

(49) *wai-wai-ʻa*
RDP-fresh.water-PROP
'watery; thin (of liquids)'

There are several colour terms that end in *ʻa*, but it is not always evident that the syllable represents the proprietive suffix. For example: *kakamira-ʻa* and *kakamira* 'yellow'; *ma-marawa-ʻa* 'green, greenish', cf. *marawa* 'green, unripe'; *kinakinaʻa* 'grey', possibly based on *kinakina* 'a bird'.

In all the cases where a word clearly contains the proprietive morpheme, the form of the morpheme is *-ʻa*. Geerts lists *oʻonara* 'thorny, prickly', which may be based on *oʻona* 'a fresh water shell fish'. (Compare Wayan *viroviroa* '[e.g. skin, timber] be rough, not smooth', possibly based on *viro* 'shell spp.'; see §2.2.) Note that the word for 'stony' is *hau-ʻa*, while in the cognate forms in other Malaita/Makira languages the proprietive morpheme has the form *-la*: Toqabaqita *fau-la* (see [13] in §2.1.1.1), Kwaraʻae *fau-la* ([41] above), Kwaio *fou-la* (44) and Ulawa *häu-le* ([51] below). Both *-/ʔa/ and *-/la/ are reconstructible for Proto Malaita/Makira (see §5), but it appears that the latter has been completely or nearly completely lost in ʻĀreʻāre, having been replaced by *-ʻa*.

Ulawa has two proprietive suffixes. One is listed by Ivens (1929:2) as *-ʻä*, which in proprietive forms can have the form *-ʻa* or *-ʻe*. According to Ivens, it derives adjectives and adverbs from nouns, verbs and other adjectives. The other suffix is listed as *-lä*, which Ivens (p.154) calls 'adjectival ending, added to verbs and participles'. It can have the form *-la* or *-le*.

ULAWA

(50) *mwaamwaa-ʻa*
catkin-PROP
'possessing catkins' (Ivens 1929:174)

(51) *häu-le*
rock/stone-PROP
'stony' (Ivens 1929:95)

Besides *mwaamwaa-ʻa* 'possessing catkins', there is also *mwaamwaa-la* 'maggoty, infested with worms', based on *mwaamwaa* 'maggot, worm' (cf. also *mwaa* 'snake'). It is not clear whether the words for 'catkin' and 'maggot, worm' are homophones, or whether this is a case of polysemy, with the two senses selecting different variants of the proprietive morpheme. (Ivens lists the two *mwaamwaa* forms as separate headwords.)

What is unusual about Ulawa is that the proprietive morpheme can occur twice in a derived form, either with the *ʻa*/*-ʻe* variants occurring twice or combining with the *-la*/*-le* variants, in that order. Ivens (1929) lists the forms *-ʻaʻa* and *-ʻala*. The former is listed together with *-ʻä*; the latter is listed by itself as 'adjectival ending added to nouns' (p.17).

ULAWA

(52) *wäi-wei-ʻaʻa*? or: *wäi-wei-ʻa-ʻa*?
RDP-fresh.water-PROP RDP-fresh.water-PROP-PROP
'watery' 'watery' (Ivens 1929:2)

(53) *sasu-ʻala*? or: *sasu-ʻa-la*?
smoke-PROP smoke-PROP-PROP
‘smoky’ ‘smoky’ (Ivens 1929:17)

While doubling of the proprietive suffix in Ulawa may be unique (but see [43] from Kwaraʻae), it is noteworthy that Ulawa also permits doubling of the nominalising suffixes (Lichtenberk, forthcoming), an uncommon kind of construction in Oceanic. However, doubling of the nominaliser is also possible in Toqabaqita, but doubling of the proprietive marker is not.

Arosi has the suffix -*ʻa* ‘forming adjectives from noun[s]’ (Fox 1978:1).

AROSI
(54) *wai-wai-ʻa*
RDP-water-PROP
‘watery’

Longgu has -*ʻa*, which is ‘a productive means of deriving verbs from nouns’ (Hill 1992:116).

LONGGU
(55) *garugaru-ʻa*
scabies-PROP
‘full of scabies’

(56) *gale-ʻa*
child-PROP
‘pregnant’ (Hill 1992:116)

Note that in (56) the proprietive suffix does not signify multitude, abundance of entities designated by the base noun.

The same suffix is probably also found in some colour terms, which are adjectives; for example, *bulubuluʻa* ‘black’ (Hill 1992:78) (cf. Toqabaqita *buu-bulu-qa* and *bulu-bulu-qa* ‘be black; be dark-coloured: dark grey, dark blue, dark purple’ in [32] in §2.1.1.2).

Nggela has a proprietive suffix -*ga*, which ‘[m]akes adjectives from nouns, with the sense of possessing the quality or thing’ (Fox 1955:16).

NGGELA
(57) *ahu-ga*
smoke-PROP
‘smoky, full of smoke’ (Fox 1955:8)

(58) *rau-rau-ga*
RDP-leaf-PROP
‘leafy’ (Fox 1955:175)

However, Fox (1955:122) also lists the suffix -*na*, to which he assigns the following functions, among others: ‘adj[ectival] suffix’ and ‘suffix forming past part[iciples]’. As an example of the former he gives the form in (59), and as an example of the latter the form in (60).

NGGELA

(59) *meto-na*
dirt/dirty-PROP?
'dirty' (Fox 1955:100)

(60) *nggere-nggere-na*
make.marks/scratches-PST.PTCP?
'striped' (Fox 1955:144)

Nggere-nggere-na appears to be based on an intransitive verb, and so the form is more likely to be a proprietive one in the sense the term is used here, rather than a past participle, 'have/having stripes'?.

In his sketch of the grammar of Nggela, Ivens (1937:1093) mentions two proprietive suffixes, *-ga* and *-a*. About *-ga* he says that 'it is added to nouns and verbs; its use with verbs is the more extensive and may be considered as one of the most characteristic features of the Florida [Nggela] language' (p.1093). One of the examples he gives has the 'be X-like' type of meaning.

NGGELA

(61) *beti* *tina-ga*
water/stream mother-PROP
'river-like' (Ivens 1937:1093)

The compound *beti tina* (*mbetitina* in Fox [1955:226]) 'water-mother' has the meaning 'river'.

The form *-a* is considered by Ivens (1937:1093) to be 'probably *-ga* through the loss of *g*'.

NGGELA

(62) *oto-oto-a*
RDP-branch-PROP
'with branches' (Ivens 1937:1093)

(63) *rau-ravu-a*
RDP?-ashes-PROP
'grey', 'pale and wan after sickness' (Fox 1955)

As Geraghty (1983:265) points out, Nggela *rauravua* is cognate with (Standard) Fijian *dravudravua*. The Fijian proprietive suffix *-a* is discussed in §2.2.

Bugotu has *-ga*, 'an adjectival suffix added to nouns' (Ivens 1940:11):

BUGOTU

(64) *bea-bea-ga*
RDP-water/liquid/juice-PROP
'watery, tasteless, insipid; without salt' (Ivens 1940:5)

(65) *faafata-ga*
layer/kind/rank-PROP
'in tiers' (cf. *fata* 'tier') (Ivens 1940:10)

Following is a non-exhaustive list of other Southeast Solomonic languages in which proprietive forms are found. In Tolo, the proprietive suffix has the form *-ha* (Crowley 1986): *kolo-ha* 'watery' (*kolo* 'water', 'river'). Inakona has *-ga*: *su-suli-ga* 'strong, able-bodied' (*suli* 'body') (Capell 1930:121). Vaturanga has *-ha*, but Ivens (1934:363) also

mentions *-a*, although the examples of the latter are not quite clear. The form *-ha* can be added both to nouns and to verbs. An example of the latter is *matahu-ha* 'fearful' (*matahu* 'to fear') (Ivens 1934:363). Although Ivens glosses *matahu-ha* with an adjective, he says that such adjectival forms 'are really verbs' (p.363). Wango has '*a*. The example given by Codrington (1885:509) is *buruburu'a* 'black'; cf. Toqabaqita *bulu-bulu-qa* 'be black; be dark-coloured: dark grey, dark blue, dark purple' in (32) in §2.1.1.2. And in Fagani, according to Codrington (1885:502), '[t]he termination *ga* is characteristic' with adjectives. As examples Codrington gives two colour terms, *purupuruga* 'black' (cf. Toqabaqita *bulu-bulu-qa*), and *merameraga* 'red' (cf. Toqabaqita *mela-mela-qa* 'be terracotta/brick coloured' [*mela* 'sp. of sea cucumber (brown in colour), also used in certain collocations usually to express brown or reddish colour']).

2.2 Proprietives elsewhere in Central/Eastern Oceanic

Proprietive forms are found in a number of languages spoken in Vanuatu. Tamambo has the suffix *-ha*, which 'derives intransitive verbs which are NOUN-like, and appears to be a fully productive process' (Jauncey 1997:139). In Tamambo, proprietive forms can express the meaning 'have, contain X' or 'be X-like'; see (66) and (67), respectively:

TAMAMBO

(66) *vulu-vulu-ha*
RDP-hair-PROP
'be hairy'

(67) *dondo-ha*
night-PROP
'be dark/very cloudy' (Jauncey 1997:139)

For another example of the meaning 'have, contain X' see (3) in §1.

The suffix *-ha* also serves to derive 'a small set of colour terms which function as adjectives' (Jauncey 1997:139); for example, *dae-ha* 'red' (*dae* 'blood'), *henja-ha* 'blue-green' (*henja* 'blue-green fish'). It is also most likely historically present in *vuriha* 'black/dirty' (Jauncey 1997:276).

Mota has a number of suffixes that Codrington (1885:273) calls 'adjectival terminations': *-ga*, *-ra*, *-ta*, *-sa* and *-la*. According to Codrington and Palmer (1896:xviii), *-sa* and *-la* are rare. Besides *-ga* there is also *-iga*, but the conditions on the use of the two variants are not clear. One example of each suffix is given in (68)–(72). In Mota both types of meaning, 'have, contain X' and 'be X-like', can be expressed by proprietive forms.

MOTA

(68) *vatu-ga*
stone/rock-PROP
'stony' (Codrington and Palmer 1896)

(69) *tapera-ta*
dish-PROP
'dish-shaped' (Codrington 1885:273)
'shallow, like a *tapera*' (Codrington and Palmer 1896:198)

(70) *lig-ligi-ra*
-PROP
'fluid, juicy' (Codrington and Palmer 1896:49)

Compare *ligiu* 'moisture, juice, sap, gravy, liquid contained'.

MOTA
(71) *vula-sa*
moon/white-PROP
'fair, in complexion' (Codrington and Palmer 1896)

Besides *vula-sa* there is also *vula-vula-sa* 'white, unripe, of a yam'.

MOTA
(72) *gae-la*
fibre/string-PROP
'stringy, tough' (Codrington and Palmer 1896)

Codrington (1885:168) mentions two colour adjectives that, according to him, contain the termination *-ga*, 'though the nouns to which the *ga* is suffixed are not in use': *agaga* 'white' and *turturuga* 'blue'.

Codrington (1885:167) lists a number of languages in present-day Vanuatu and the 'adjectival terminations' they have. Most of the languages have the suffix *-ga* or *-g*, and some also have *-ra* or *-r*. For Marino, called Maewo by Codrington,[4] he also lists *-sa*, besides *-ga* and *-gi*. (However, it is not clear that *-gi* is really a member of the same set with *-ga* and *-sa*. There is also *-si*, besides *-sa*, which occurs in transitive verbs.) Two examples from Marino are given in (73) and (74). As (73) shows, in Marino a proprietive form can have the meaning 'be X-like'.

MARINO
(73) *angoa-ga*
turmeric-PROP
'yellow'

(74) *tangtangi-sa*
?-PROP
'merciful', 'pitiful' (Codrington 1885:412)

Codrington gives the form of the noun 'turmeric' as *ango*. The presence of the middle *a* in *angoa-ga* is not explained. He does not give the meaning of the base from which *tangtangi-sa* is derived, but it most likely reflects POc *tangis 'cry, weep', with reduplication.

Example (75) from Arag (Raga) shows a proprietive form with the suffix *-g*. The word is cognate with Marino *angoa-ga* 'yellow', but without the middle *a*:

ARAG (RAGA)
(75) *ango-ga*
turmeric-PROP
'yellow' (Codrington 1885:435)

---

[4] As one of the referees points out, there are three languages spoken on Maewo Island. The language discussed by Codrington is spoken at the northern end of the island, which is the area where Marino is spoken.

Proprietive forms are also found in Fijian and Polynesian languages. There the suffix typically, although not exclusively, has the form *-a*. According to Geraghty (1983:265), *-a* 'makes statives out of (usually) reduplicated nominals throughout Fiji'. Churchward (1941:45) also points out that the root is usually reduplicated in the proprietive forms in (Standard) Fijian. Capell (1973:1) considers Fijian *-a* to be 'a participial ending used in forming adjs. from some nouns'. The proprietive forms can have either of the two types of meaning found in other languages, 'have, contain X' and 'be X-like'.

(STANDARD) FIJIAN

(76) *ulo-ulo-a*
RDP-worm/maggot-PROP
'wormy, maggoty' (Capell 1973:247)

(77) *dravu-dravu-a*
RDP-ashes-PROP
'covered with ashes, grey in colour', 'poor, poverty' (Capell 1973)

In Boumaa Fijian, *dravu-dravu-a* '(related to the noun *dravu* 'ashes') means "poor, poverty-stricken" when describing a person, but "grey" when used of an animal or thing' (Dixon 1988:231).

Both in (Standard) Fijian and in Boumaa, proprietive forms can be based on verbs. The example in (78) is from Fijian:

(STANDARD) FIJIAN

(78) *vora-vorā* (from *vora-vora-a*)
RDP-resist.command:PROP
'violent, tyrannical' (Capell 1973)

Wayan has the suffix *-a*, '[n]on-productive [verbal] suffix occurring in a few stative verbs, indicating abundance of a thing' (Pawley and Sayaba 2003). For example:

WAYAN FIJIAN

(79) *tubu-a*
TUBU-PROP
'(of a garden, etc.) be overgrown with weeds'

The base *tubu* functions both as a verb and as a noun with a variety of meanings, including 'sprout', 'grow' and 'new growth, young shoot or sprout'.

There appears to be at least one proprietive form in Wayan with the 'be X-like' type of meaning: *viroviroa* '(of surface, e.g. skin, timber) be rough, not smooth'. There is a noun *viro* 'shellfish … taxon: generic for Augers (*Terebra* spp.) and Miters (*Mitra* spp.)'. *Viro* shells are pointy and rough, and so it is possible that *viroviroa* is indeed based on *viro* (Andrew Pawley, pers. comm. 20 May 2008). Compare 'Āre'āre *o'onara* 'thorny, prickly', possibly based on *o'ona* 'a fresh water shell fish' (Geerts 1970; see §2.1.2 above).

There may be at least one proprietive form in Fijian where the 'have, contain X' type of meaning does not necessarily signify an abundance of X: *luve-a*, based on *luve* 'offspring'. Churchward (1941:45) glosses it as 'having offspring', and Capell (1973:129) as 'having children'. Capell also says that *luvea* is used chiefly in the expression *luvea vakalevu* 'having a large family' (*vaka-* intensifying prefix, *levu* 'big, great, large').

Churchward (1941:46) points out that '[i]n some cases the suffixed *a* is preceded by a consonant'. For example:

(STANDARD) FIJIAN
(80) *drega-drega-ta*
RDP-gum-PROP
'sticky'

(Capell [1973] gives the meaning of *drega* as 'gum or glue which exudes from trees or fruits, gum, paste, glue [in general]'.)

Drega-drega-ta has the 'be X-like' type of meaning.

Churchward (1941:46) also gives *qeleqelewa* 'soiled, dirty', based on *qele* 'soil, dirt'. However, Capell (1973:163) gives the derived form as *qeleqelea* 'dirty', figuratively 'valueless', without the *w* (*qele* 'earth, soil'). Churchward also suggests (p.46) that *sokosokota* 'thick (of liquids)', *kamikamica* 'sweet' and *lumilumisa* 'shine' contain suffixes with consonants, *-ta*, *-ca* and *-sa*, respectively, but without identifying the bases from which the forms would be derived. Capell does not list *sokosokota*; he does list *kamikamica* 'sweet, agreeable, well tasting, of food, also figuratively of a person, speech, etc.', but no *kami*; and instead of *lumilumisa* he gives *lumilumisā* (also *lumisā*) 'shining, resplendent', based on *lumi* 'moss', 'kind of edible seaweed' (p.127).

In Polynesian languages, the proprietive suffix has the form *-a*. Examples from Pukapukan and Samoan appear, respectively, in (1) and (2) in §1. The example in (81) below comes from Tongan:

TONGAN
(81) *efu-a*
dust-PROP
'covered with dust, dusty' (Churchward 1985:244)

In Tokelauan, *-a* 'is a very productive suffix which is added to nouns to form verbs which mean having an abundant supply of that noun' (*Tokelau dictionary* 1986:xxxvi-xxxvii). For example:

TOKELAUAN
(82) *namu-a*
mosquito-PROP
'be infested with mosquitoes'

For Proto Polynesian, POLLEX (Biggs and Clark 2000) gives *-a 'presence of, or infestation by, that indicated by the noun'.

3 Proprietives in Western Oceanic

Proprietive forms are also found in some Western Oceanic languages. Motu has *-ka*, which is 'suffixed to the adjective [and] always intensifies the quality' (Lister-Turner and Clark n.d.a:32). The typical pattern is for the *-ka*-suffixed form to be fully reduplicated:

MOTU
(83) *kuro-ka-kuro-ka*
whitish-PROP-whitish-PROP
'dazzling white' (Lister-Turner and Clark n.d.a:32)

Compare *kurokuro* 'white'.

MOTU

(84) *duhu-ka-duhu-ka*
muddy-PROP-muddy-PROP
'very muddy' (Lister-Turner and Clark n.d.a:32)

Compare *duhuduhu* 'slightly cloudy'.

However, in some cases *-ka* is added to nouns. The derived form may have the 'have, contain X' or the 'be X-like' type of meaning; see (85) and (86), respectively:

MOTU

(85) *ranu-ka-ranu-ka*
water/juice/liquid-PROP-water/juice/liquid-PROP
'watery' (Lister-Turner and Clark n.d.b)

(86) *era-ka-era-ka*
turtle-PROP-turtle-PROP
'broad' (used with *keme* 'chest') (Lister-Turner and Clark n.d.b)

The form in (87) has both types of meaning, 'have, contain X' and 'be X-like':

MOTU

(87) *rara-ka-rara-ka*
blood-PROP-blood-PROP
'bloody', also 'rusty' (Lister-Turner and Clark n.d.b)

There is at least one case where the *-ka*-suffixed form does not undergo reduplication:

MOTU

(88) *tadi-ka*
seawater-PROP
'salty' (Lister-Turner and Clark n.d.b)

Mekeo, a close relative of Motu, too has a proprietive suffix *-ka* (/ka/; /ʔa/ in East Mekeo):

MEKEO

(89) *ini-ka*
thorn-PROP
'thorny'

(90) *vei-ka*
water-PROP
'wet, watery'

(91) *pui-ka*
hair/bristle-PROP
'hairy', 'bristly' (Jones 1998:136)

In Mekeo the proprietive suffix is not productive. Jones (1998:136) says that *-ka* is one of a small number of suffixes that 'appear to make no distinct contribution to the semantics of the root but merely mark it overtly as "adjectival" [footnote omitted]'. Specifically about *-ka* Jones says that '… it seems to intensify the root. In some cases, however, it signifies "having, being, characterised by" [footnote omitted]. It may in some cases, however, represent a shortened form of *kae* "ascend, upward, up"' (p.136). As we have seen, cognates of the

proprietive suffix are found elsewhere in Oceanic, and there is no reason to assume that Mekeo proprietive *-ka* comes from *kae*.

Proprietive forms are also found in Vitu, where the suffix has the form *-a*. The suffix derives adjectival/stative verbs (van den Berg and Bachet 2006:46). Some examples are given below (all of them from page 48):

VITU
(92) *habu-habu-a*
fog-PROP
'foggy'

(93) *muku-muku-a*
RDP-sore-PROP
'full of sores'

In nearly all of the examples given by van den Berg and Bachet, the base is a noun. They do list one proprietive form whose base appears to be an (adjectival) verb:

VITU
(94) *puzo-puzo-a*
RDP-white-PROP
'white; yellow (of coconut)'

As in other languages, a proprietive form can have either type of meaning; compare (92) and (93) above on the one hand and (95) below on the other.

VITU
(95) *hango-hango-a*
RDP-yellow.ginger.like.plant-PROP
'yellow'

Van den Berg and Bachet (2006:48) point out that many colour terms are formed in this way. However, '[i]n a number of cases, the root does not exist by itself'; for example: *here-here-a* 'red' and *kobo-kobo-a* 'green'.

Finally, Takia has a suffix *-a* that derives adjectives. The base may be a noun, an intransitive verb or an adjective (Malcolm Ross, pers. comm. 1 April 2008; see also Ross 2002a). When the base itself is an adjective, it is not clear what the semantic difference is between the base and the derived form.

TAKIA
(96) *dagom-a-n*
peace/tameness-PROP-3SG.POSS
'peaceful' (Ross 2002a:226)

(The third person singular possessive suffix occurs in the citation forms of most adjectives.)

The next set of examples is courtesy of Malcolm Ross (pers. comm. 1 April 2008):

TAKIA
(97) *gan-a*
be.damp/rotten-PROP
'damp, rotten'

(98) *kawil-a*
clever-PROP
'clever'

A proprietive form can have the 'be X-like' type of meaning:

TAKIA

(99) *katuk-a*
stump-PROP
'short'
(Note: *Katuk* 'stump' is a noun.)

4 Proprietives in Mussau

The Mussau data come from Ross (2000:332). According to Ross, the proprietive suffix occurs on some adjectives whose base is a reduplication of the root. The proprietive suffix has the form '*-e* after a high vowel and *-a* elsewhere'. The suffix *-na* 'is the third singular cross-referencing marker and occurs on certain adjectives'. In all three cases where the root can be identified, it is a noun. And in all three cases the meanings are of the 'be X-like' type.

MUSSAU

(100) *bo-boŋi-e-na*
'black' (cf. *bo* 'night')

(101) *riu-riu-e-na*
'thin (of animates)' (cf. *riu* 'bone')

(102) *rae-rae-a-na*
'red' (cf. *rae* 'blood')

The proprietive suffix is also presumably present in *vero-veroŋ-a-na* 'black', even though there is no evidence of a corresponding root.

5 Proprietives in Proto Oceanic and in Oceanic languages: summing up

In the preceding sections the focus has been on those proprietive affix forms that can be reconstructed for POc and those that are in a paradigmatic relation with the reflexes of the POc forms; see Table 1 below. Some languages have other ways of forming proprietive words. For example, in Jabêm '[m]any adjectives are derived from nouns by prefixing the ligature /ŋa-/ ….' (Ross 2002b:279). The three examples where the internal structure is clear are all proprietive, including /ŋa-lɛmɔŋ/ 'muddy, soft' (/lɛmɔŋ/ 'mud'). And in Kosraean, proprietive adjectives are formed by reduplication of nouns; for example *fohk-fohk* 'dirty' (*fohk* 'dirt') (Lee 1975:225). Ross (2000) also mentions reduplication as a way to form proprietive lexemes, with special reference to colour terms. For reduplication in Oceanic colour terms see also Blust (2001). As exemplified in §2–§4, reduplication of the base and the proprietive suffix often cooccur, but either strategy can form proprietive lexemes by itself.

The proprietive forms discussed or mentioned in the preceding sections are listed in Table 1. The first column contains reflexes of *-ka, and the second column reflexes of *-a. A few languages contain other variants of the proprietive morpheme. These are listed in the third column. It will be argued further below that some of the latter are descendants, with a secondary development, of *-a.

Table 1 shows there is evidence for reconstructing both *-ka and *-a as the proprietive suffixes in POc.

Pawley (1972) reconstructed *-ka and *-a for what he called 'Proto Eastern Oceanic'. He characterised the function of *-ka ~ *-a as 'stative verb derivative, denoting an abundance of. Transforms nouns into stative verbs, as *pulupula-ka [sic] '(to be) hairy', from *pulupulu 'body hair'' (p.39). Elsewhere (p.83) he called the derived forms 'adjectival verbs'. Pawley offered no comments on the conditions governing the use of the two variants *-ka and *-a.

Lynch et al. (2002:75) mention *-[k]a for POc, which 'seems to have derived adjectival nouns from other roots, usually nouns (Ross 1997b), *-ka occurring on vowel-final roots, *-a on consonant-final', although no evidence is provided concerning the distribution of the two variants.[5] This echoes Ross' (2000) reconstruction and discussion of POc *-[k]a, mentioned in §1 above. There Ross attributes to *-[k]a the function of deriving adjectival nouns. According to Ross, POc had only two open classes of lexemes: nouns and verbs. Lexemes that designated properties were either adjectival nouns or adjectival verbs. The subclass of adjectival nouns, which included proprietives, was small and closed, while the subclass of adjectival verbs was large. Ross (2000:332) suggests that the distribution of the two allomorphs of the proprietive suffix *-ka and *-a was conditioned phonologically: 'In POc *-ka seems to have occurred on vowel-final bases, *-a on consonant-final bases.'

Table 1: Proprietive morphemes in Oceanic languages

CENTRAL/EASTERN OCEANIC			
Toqabaqita	-qa (/ʔa/)		-la
Lau	-a[a]		-la, -ra, -na?
Kwara'ae	-'a		-la
Kwaio	-'a		-la
'Āre'āre	-'a		
Ulawa	-'a,[b] -'e[c]		-la,[d] -le[e]
Arosi	-'a		
Longgu	-'a		
Nggela	-ga (/ɣa/)	-a	
Bugotu	-ga (/ɣa/)		
Tolo	-ha		
Inakona	-ga (/ɣa/)		
Vaturanga	-ha	-a?	
Wango	-'a		
Fagani	-ga		
Tamambo	-ha		
Mota	-ga (/ɣa/)		-ra, -ta, -sa, -la
Marino	-ga		-sa, -gi?
Arag/Raga	-ga (/ɣa/)		
(Standard) Fijian		-a	-ta
Boumaa Fijian		-a	
Wayan		-a	
Pukapukan		-a	

[5] I have not been able to consult Malcolm Ross' unpublished manuscript dated 1997b in the quote.

Samoan		-a
Tongan		-a
Tokelauan		-a
WESTERN OCEANIC		
Motu	-ka	
Mekeo	-ka, -ʔa	
Vitu		-a
Takia		-a
MUSSAU[f]	-a/-e?	-a/-e?

[a] In Lau, /ʔ/ is lost in certain positions. *-a* is considered here to be a reflex of *-ka, rather than of *-a.
[b] Can be doubled as *-ʻaʻa*, and can combine with *-la* as *-ʻala*.
[c] *-ʻe* is also written as *-ʻë*.
[d] Can combine with *-ʻa* as *-ʻala*.
[e] *-le* is also written as *-lë*.
[f] POc *k is variably lost in Mussau (Blust 1984), and so the proprietive morpheme could reflect either *-ka or *-a. For that reason the Mussau reflexes are listed in both columns.

The kind of phonological conditioning posited by Lynch et al. is plausible on general phonological grounds, because POc did not permit sequences of consonants. However, the evidence for it is not sufficient. This is because the variant *-a could have occurred not only after consonant-final but also after vowel-final bases, with the distribution of *-ka and *-a perhaps conditioned lexically in the latter type of case. Nggela has reflexes of both variants, and *-a* does occur with vowel-final stems: *oto-oto-a* ‘with branches’ (see [62] in §2.1.2) and *rau-ravu-a* ‘grey’, ‘pale and wan after sickness’ (63).

Even though POc had two variants of the proprietive morpheme, *-ka and *-a, the usual pattern in the modern languages is to retain the reflex of only one of them, either exclusively or nearly so. Thus, the languages that belong in the Southeast Solomonic subgroup have, nearly exclusively, reflexes of *-ka (apart from reflexes of the innovative *-la variant in one subgroup), while the Polynesian languages have exclusively reflexes of *-a, and Fijian nearly so. For this reason it is difficult to determine the distribution of the two POc allomorphs. (I am grateful to Bethwyn Evans for reminding me of that [pers. comm. 5 February 2009].)

As shown in Table 1 and in the relevant places in §2.2, a few languages have proprietive forms that are not reflexes of *-ka and are not, straightforwardly, reflexes of *-a: *-ra*, *-ta*, *-sa*, *-la* in Mota; *-sa* in Marino; and *-ta* in Fijian. There are also *l*-initial forms in languages that belong in the Malaita/Makira subgroup, a low-order grouping within Southeast Solomonic: *-la* in Toqabaqita, Lau, Kwaraʻae, Kwaio and *-la*/*-l*e in Ulawa. These appear to be the result of an independent development and will not be considered in this context; but see further below.

There is evidence that in at least some cases, the initial consonants in the proprietive suffixes in Fijian and in Marino reflect earlier stem-final consonants that have been reanalysed as part of the proprietive morpheme when word-final consonants became disallowed in the histories of these languages. (Compare the reanalysis of verb-final consonants as parts of transitive suffixes in some Oceanic languages.) Bases that ended in a consonant took the *-a variant of the proprietive morpheme. In the languages in question, the stem final consonant was dropped in the absence of a suffix but was retained in the presence of the suffix *-a. Consider (Standard) Fijian *drega-drega-ta* ‘sticky’, based on the

noun *drega* ‘gum’, where the proprietive suffix has the form *-ta* (see [80] in §2.2). There is a related transitive verb *drega-ta* ‘glue, stick together’ (Capell 1973), where *-ta* is a transitive suffix, with the same thematic consonant. (The proprietive suffix *-ta* and the transitive suffix *-ta* are different morphemes.) And in Marino, the *s* of the suffix in *tangtangi-sa* ‘merciful’, ‘pitiful’ (see [74] in §2.2) evidently comes from the final consonant in POc *tangis ‘cry, weep’.

However, there are also cases whose history is unclear. I will give one illustrative example. If POc had a proprietive form for ‘be mossy, be covered with moss’, one would expect it to be *lumut-a and/or *limut-a; cf. *lumut and *limut ‘generic term for mosses, algaes and seaweeds’ (Evans 2008:77). Now, Mota has *lumta-ga* (*lumuta-ga*) ‘covered with moss’, based on *lumuta* ‘moss’ (Codrington and Palmer 1896:56). The erstwhile final *t* was protected from the process of final-consonant loss through the addition of *a*. The reflex of the POc proprietive form *-ka was added to the base *lumuta. It is not clear whether *lumuta was monomorphemic with the final *a* protecting the final *t* from deletion, or whether it was in fact the reflex of the POc proprietive form *-a. Under the latter scenario, pre-Mota *lumuta had the meaning ‘mossy, covered with moss’, was later reinterpreted as a noun ‘moss’, to which then the *-ga* reflex of POc *-ka was added: *lumuta-ga* ‘covered with moss’. A different historical process has taken place in Toqabaqita. Toqabaqita has a stative verb *lumu-qa* ‘be mossy; be covered with moss, lichen’, based on *lumu* ‘moss, lichen’.[6] The developments in the histories of Mota and Toqabaqita suggest that the proprietive morpheme remained productive. In Mota it was added either after the supporting final *a* was added or after the final *-a*, originally proprietive, ceased to be interpreted as such. And in the history of Toqabaqita it was added after the loss of the final consonant in *lumut.

In some languages the proprietive derivation is productive, or, at least, there is a relatively large number of proprietive forms (e.g. in Tamambo and in Toqabaqita), while in others the number of proprietive forms is small and the derivational process is not productive (e.g. in Wayan and in Mekeo). The relatively high productivity of proprietive derivation in various languages makes it difficult to determine whether cognate proprietive forms reflect a POc proprietive form or whether they are the result of independent parallel developments in daughter (proto)languages, a point also made by Ross (2000:334). For example, Ross (2000:336) has reconstructed *[yaŋo]yaŋo ‘yellow’ for POc, based on *yaŋo ‘turmeric, *Curcuma longa*’. Forms for ‘yellow’ that involve reduplication are found in different primary subgroups of Oceanic. Ross also lists Raga *aŋo-ɣa* ‘yellow’ with the proprietive suffix. He also lists Vitu *ɣaŋo-ɣaŋo* ‘yellow’. However, van den Berg and Bachet (2006) give *hango-hango-a*[7] for ‘yellow’, with the proprietive suffix, based on *hango* ‘yellow ginger-like plant’. Raga and Vitu belong in different primary subgroup of Oceanic. Should one then reconstruct a form for ‘yellow’ with the proprietive suffix for POc, or are the Raga and the Vitu forms results of independent developments? Note that Raga reflects the *-ka variant of the POc proprietive morpheme, while Vitu reflects the *-a variant.

The following example too can be seen as possible evidence of independent developments of a proprietive etymon. POc had *waiR ‘fresh water; river, stream’ (Osmond et al. 2003:57). If there was a corresponding proprietive form in POc, one would

---

[6] Besides *lumu-qa* ‘be mossy; be covered with moss, lichen’ Toqabaqita also has synonymous forms *lumu-lumu-qa*, *luu-lumu-qa, rumu-qa*, *rumu-rumu-qa* and *ruu-rumu-qa*, based, respectively, on *lumu-lumu*, *luu-lumu, rumu*, *rumu-rumu* and *ruu-rumu*, all of which are synonymous with *lumu* ‘moss, lichen’.

[7] Orthographic *h* represents /ɣ/, and *ng* represents /ŋ/.

expect it to be *waiR-a ‘watery’. Samoan has *vai-a* ‘watery’ (see [2] in §1), which would be a regular reflex of *waiR-a. However, other languages have cognate proprietive forms that contain a reflex of the *-ka variant of the proprietive suffix: Mekeo *vei-ka* ‘wet, watery’ ([90] in §3); and ‘Āre‘āre has *wai-wai-*‘a ‘watery; thin (of liquids)’ ([49] in §2.1.2), and Arosi has *wai-wai-*‘a ‘watery’ ([54] in §2.1.2). The Mekeo, ‘Āre‘āre and Arosi forms are later developments after the loss of the R in *waiR. In Polynesian, there has been wholesale replacement of the *-ka variant by the *-a variant, and one cannot tell whether Samoan *vai-a* ‘watery’ directly reflects POc *waiR-a or it too is a later development.

In the ancestor of Toqabaqita, Lau, Kwara‘ae, Kwaio and Ulawa, a new variant of the proprietive morpheme developed, *-la. Dissimilation most likely played a role: *-/ʔa/ ceased to be used with bases that had /ʔ/ in their final syllable, where *-/la/ was used instead.

Lynch et al. (2002:75) say that ‘*-[k]a seems to have derived adjectival nouns’, adjectival nouns being ‘a small, closed class of attributive postmodifiers’ (p.74), that is, postmodifiers of nouns. (See also Ross 2000.) However, in the modern languages the proprietive forms typically function as stative verbs or as adjectives. Grammars do not always make a principled distinction between the two classes of lexemes, but the derived forms are (normally) not nouns. On the basis of the present-day evidence the conclusion is that POc proprietive forms were stative verbs.

In most cases the proprietive lexemes are derived from nouns, the noun often being reduplicated in the proprietive form. The derived forms have one of two types of meaning, ‘have, contain X’ or ‘be X-like’, or possibly both, where X represents the kind of entity, stuff, etc. designated by the noun of the base. The ‘have, contain X’ type of meaning usually carries the implication of an abundance, large quantity/amount of X, but not necessarily so. Nevertheless, in the latter type of case the property in question is significant with respect to the relevant participant, for example, ‘having a child/children’ (that is, not being childless [any longer]).

As far as the second type of meaning is concerned, ‘be X-like’, it is found in many languages with a number of proprietive lexemes that designate colours; for example ‘be blood-like’, i.e. ‘be red’. This was also presumably the case in POc, as also pointed out in Lynch et al. (2002).

As far as POc is concerned, one can conclude that it had proprietive forms based on nouns, and they could have either type of meaning.

In some languages proprietive forms can be based on stative verbs, as well as on nouns. This may have also been the case in POc. It is not always clear what the difference in meaning is between the base stative verb and the derived proprietive form, but in some cases the meaning of the derived form is intensive.

Finally, one of the referees asks whether the nouns that serve as bases in proprietive forms can have their own modifiers, such as ‘(with) black hair-prop’. There is no evidence of such structures in any of the Oceanic languages for which information on proprietives is available, with one possible, partial exception. Toqabaqita has teqe-bali-qa ‘be asymmetrical, lopsided’, based on teqe ‘one’ and bali ‘side’ (see [17] in §2.1.1.1), but the semantics is irregular. (The meaning is not ‘be one-sided; have one side’.) In other words, the process forming proprietive forms is purely derivational. It forms adjectives or stative verbs from nouns or stative verbs, and such derived forms do not exhibit properties that make them syntactically different from other adjectives or stative verbs in the language.

References

Biggs, Bruce and Ross Clark, eds. 2000. POLLEX (The Comparative Polynesian Lexicon Project). Electronic manuscripts. Auckland: University of Auckland.

Blust, Robert A. 1984. A Mussau vocabulary, with phonological notes, 159–208. Canberra: Pacific Linguistics.

—— 2001. Reduplicated colour 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.

Capell, A. 1930. The language of Inakona, Guadalcanar, Solomon Islands. *Journal of the Polynesian Society* 39.113–136.

—— 1973. *A new Fijian dictionary* (4th ed.). Suva: Government Press.

Churchward, C. Maxwell. 1941. *A new Fijian grammar*. Suva: Government Press.

—— 1985 [1953]. *Tongan grammar*. Tonga: Vava‘u Press.

Codrington, R.H. 1885. *The Melanesian languages*. Oxford: Clarendon Press.

Codrington R.H. and J. Palmer. 1896. *A dictionary of the language of Mota, Sugarloaf Island, Banks’ Islands*. London: Society for Promoting Christian Knowledge.

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.

Dixon, R.M.W. 1988. *A grammar of Boumaa Fijian*. Chicago/London: University of Chicago Press.

Evans, Bethwyn. 2008. Ethnobotanical classification. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society,* vol. 3: *Plants*, 53–84. Canberra: Pacific Linguistics.

Fox, Charles E. 1955. *A dictionary of the Nggela language (Florida, British Solomon Islands)*. Auckland: Unity Press.

—— 1974. *Lau dictionary*. Canberra: Pacific Linguistics.

—— 1978. *Arosi dictionary*. (Revised ed. with English–Arosi index prepared by Mary Craft.) Canberra: Pacific Linguistics.

Geerts, P. 1970. *‘Āre‘āre dictionary*. Canberra: Pacific Linguistics.

Geraghty, Paul A. 1983. *The history of the Fijian languages*. Honolulu: University of Hawai‘i Press.

Hill, Deborah. 1992. Longgu grammar. PhD thesis. Canberra: The Australian National University.

Ivens, Walter G. 1929. *A dictionary of the language of Sa‘a (Mala) and Ulawa, South-east Solomon Islands*. London: Oxford University Press; Melbourne: Melbourne University Press.

—— 1934. A grammar of the language of Vaturanga, Guadalcanal, British Solomon Islands. *Bulletin of the School of Oriental Studies* 7.349–375.

—— 1937. A grammar of the language of Florida, British Solomon Islands. *Bulletin of the School of Oriental Studies* 8.1075–1110.

—— 1940. *A dictionary of the language of Bugotu, Santa Isabel Island, Solomon Islands*. London: Royal Asiatic Society.

Jauncey, Dorothy. 1997. A grammar of Tamambo, the language of western Malo, Vanuatu. PhD thesis. Canberra: The Australian National University.

Jones, A.A. 1998. *Towards a lexicogrammar of Mekeo (an Austronesian language of West Central Papua)*. Canberra: Pacific Linguistics.

Keesing, Roger M. 1985. *Kwaio grammar*. Canberra: Pacific Linguistics.

Lee, Kee-Dong (with the assistance of Lyndon Cornelius and Elmer Asher). 1975. *Kusaiean reference grammar*. Honolulu: University of Hawai‘i Press.

Lichtenberk, Frantisek. 2008a. *A dictionary of Toqabaqita (Solomon Islands)*. Canberra: Pacific Linguistics.

—— 2008b. *A grammar of Toqabaqita*. Berlin/New York: Mouton de Gruyter.

—— Forthcoming. Nominalizations in Toqabaqita and closely related languages. In Foong Ha Yap and Janick Wrona, eds *Nominalization in Asian languages: diachronic and typological perspectives*. Amsterdam/Philadelphia: John Benjamins.

Lister-Turner, R. and J.B. Clark. n.d.a. *A grammar of the Motu language of Papua* (2nd edition, edited by Percy Chatterton). Sydney: Government Printer.

—— n.d.b. *A dictionary of the Motu language of Papua* (2nd edition, edited by Percy Chatterton). Sydney: Government Printer.

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

Mosel, Ulrike and Even Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Scandinavian University Press.

Osmond, Meredith, Andrew Pawley and Malcolm Ross. 2003. The landscape. 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*, 35–86. Canberra: Pacific Linguistics.

Pawley, Andrew. 1972. On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol. 3, 1–142. Honolulu: Bernice Pauahi Bishop Museum.

Pawley, Andrew and Timoci Sayaba. 2003. Words of Waya: a dictionary of the Wayan dialect of the Western Fijian language. Unpublished manuscript. Canberra: The Australian National University.

Ross, Malcolm. 2000. Proto Oceanic adjectival morphology: the suffix *-[k]a. In Steven Roger Fischer and Wolfgang B. Sperlich, eds *Leo Pasifika: Proceedings of the Fourth International Conference on Oceanic Linguistics, Niue Island (South Pacific), 5th–9th July 1999*, 326–342. Auckland: Institute of Polynesian Languages and Cultures.

—— 2002a. Takia. In John Lynch, Malcolm Ross and Terry Crowley, eds, 2002, 216–248.

—— 2002b. Jabêm. In John Lynch, Malcolm Ross and Terry Crowley, eds, 2002, 270–296.

Ross, Malcolm, Andrew Pawley and Meredith Osmond. 2008. Introduction. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic: the culture and environment of ancestral Oceanic society*, vol. 3: *Plants,* 1-24. Canberra: Pacific Linguistics.

Salisbury, Mary C. 2002. A grammar of Pukapukan. PhD thesis. Auckland: University of Auckland.

Tokelau dictionary. 1986. Apia: Office of Tokelau Affairs.

van den Berg, René and Peter Bachet. 2006. *Vitu grammar sketch*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics. (Available online: http://www.sil.org/pacific/png/pubs/47977/Vitu_Grammar_Sketch_vol51.pdf. Accessed 24 June 2008.)

20 *At sixes and sevens: the development of numeral systems in Vanuatu and New Caledonia*

JOHN LYNCH

1 Introduction[1]

In a review of Lynch, Ross and Crowley's (2002) *The Oceanic languages*, Robert Blust went on a self-confessed 'major digression' (2005:556). He directed attention to 'Vanuatu and southern Melanesia [New Caledonia and the Loyalties], where Papuan languages are absent, but Papuan features such as [quinary numeral systems and serial verb constructions] are present' (554), and concluded that 'the Papuan features of language, culture, and physiognomy that are common to [Austronesian] speakers in Vanuatu and southern Melanesia must have been acquired by contact *in situ*' (555) — in other words, by contact with Papuan languages which must once have been spoken in that area but which have since presumably died out.

In this paper, I investigate the nature of the numeral systems of the languages of Vanuatu and New Caledonia. I should point out from the outset that this paper is *not* an attempt to prove Blust right or wrong. It may well be that the structural shift from the purely decimal system of Proto Oceanic to systems based on 5 in many of these languages was due to Papuan influence. However, there is not a single shred of evidence to suggest that this was the case, since there are no Papuan languages spoken in this area, and no evidence — either that I am aware of or advanced by Blust — that there ever were. Rather, what this paper does is to attempt to outline the morphological processes involved in the development of those systems which are not purely decimal. Given the explanations I propose, it *may* be possible to find extant Papuan languages in the Pacific which could form the basis for calques which arose in Vanuatu and New Caledonia.

[1] It is a pleasure to be able to offer this paper in honour of Malcolm Ross, whose contribution to Austronesian and Papuan linguistics over the last two decades have been enormous. Malcolm made a somewhat late entry into these fields, stimulated by the diverse linguistic environment during his time as an English master at Keravat National High School and later Principal of Goroka Teacher's College, both in Papua New Guinea, during the 1980s. His mid-career shift in interest, however, has benefited Austronesian and Papuan linguistic research tremendously.

I thank Isabelle Bril, Michael Dunn, Robert Early, Peter Lincoln, Claire Moyse, Åshild Næss and Hannah Vari-Bogiri for supplying data and/or interpretations of data from various languages; and Beth Evans, Alexandre François and Peter Lincoln for comments on an earlier draft. Responsibility for errors and misinterpretations is mine alone.

Following earlier studies (e.g. Tryon 1976; Clark 1985), I use the following terms to describe different kinds of systems:[2]

- DECIMAL refers to systems in which the numerals 1–10 are monomorphemic (or have monomorphemic roots with a synchronic or fossilised prefix), and where 20 is represented by a compound involving 2 and 10.
- IMPERFECT DECIMAL refers to systems which differ from decimal systems only in that the numerals 6–9 are compounds generally involving the numerals 1–4 in some way. In these systems, 10 and 20 are constructed in the same way as for decimal systems.
- QUINARY refers to systems in which there are no monomorphemic numerals above 5; 10 is either 5-(and)-5 or 'two fives/hands'; and 20 is a compound of the type 'one person'.[3]
- MIXED refers to systems with features of both imperfect decimal and quinary systems. Typically, 10 is 'two fives/hands', but 20 is 'two tens'.

This paper deals mainly with the numerals 1–10, though I will briefly mention forms for 20 and make mention in passing of higher numerals.

The ancestor of the languages of this area, Proto Oceanic, had a decimal system (see, for example, (1) below). Where imperfect decimal, quinary and mixed systems have developed, they involve the replacement of monomorphemic forms with analytical forms. I will therefore use the term INNOVATIVE to refer to imperfect decimal, quinary and mixed systems when they do not need to be distinguished from each other, and the term LIGATURE (abbreviated LIG) to refer to elements which conjoin numerals in compounds (e.g. Tolomako *lina-rave-rua* 5-LIG-2 'seven'). Arabic numerals as glosses appear without quotation marks.

2 Distribution of decimal and innovative systems

Full decimal systems are found in the following areas in Vanuatu–New Caledonia:

- all the languages of Ambae;
- Raga in the north of Pentecost;
- the Cape Cumberland area of the extreme northwest of Santo and, at the opposite end of the island, the offshore islands of Malo and its neighbours; and
- the northern part of Malakula, adjacent to the Malo area.

A sample is given in (1).[4]

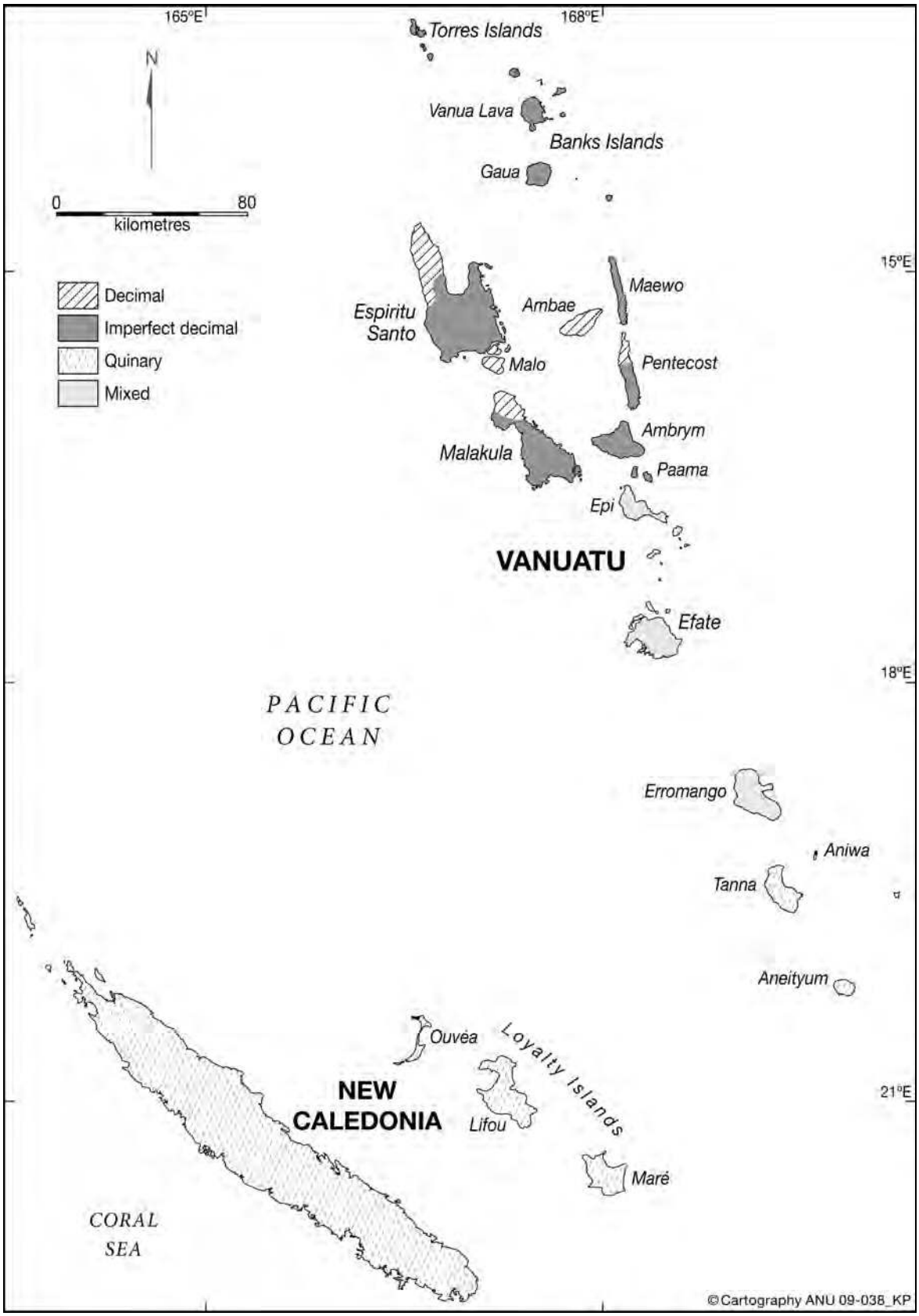
Conversely, innovative systems are found in much of Northern Vanuatu, almost all of Central Vanuatu, and throughout the two more southerly groups. Map 1 illustrates the geographical location of each type of system.

---

[2] These terms, or terms like them, were used by Tryon (1976) and Clark (1985) in their studies of Vanuatu languages. They differ slightly from terms used by, for example, Comrie (2008); but these differences are not significant.

[3] Lincoln (in press) notes that so-called quinary systems are not really quinary or base–5 in the mathematical sense, since the 'milestones' are not 5, 52 = 25, 53 = 125, etc. However, this term has been so widely used in Oceanic studies that I will retain it here.

[4] Sources of language data and comments on the orthography used can be found in the appendix. POc - Proto Oceanic; PEOc - Proto Eastern Oceanic; PNCV - Proto North-Central Vanuatu.



Map 1: Geographical location of numeral systems in Vanuatu and New Caledonia

(1)	**POc**	**Raga**	**Araki**	**Uripiv**
		Pentecost	Santo	Malakula
1	various[5]	*tea*	*hese*	*san(san)*
2	*rua	*rua*	*dua*	*e/ru*
3	*tolu	*tolu*	*rolu*	*i/tul*
4	*pat(i)	*vasi*	*v'ari*	*i/vij*
5	*lima	*lima*	*lim'a*	*e/lim*
6	*onom	*ono*	*hai/ono*	*ou/won*
7	*pitu	*bitu*	*hai/p'iru*	*e/but*
8	*walu	*v^{w}elu*	*ha/ualu*	*o/wil*
9	*siwa	*siva*	*hai/sua*	*e/siw*
10	*sa-[ŋa]-puluq	*haŋvul(u)*	*saŋavulu*	*esŋavəl*
20	*rua-ŋa-puluq	*ŋavul-gai-rua*	*ŋavul dua*	*ŋavəl eru*

2.1 Innovative systems in Vanuatu and mainland New Caledonia

Innovative systems in Vanuatu and New Caledonia (excluding the Loyalties, discussed in the next section) can be subdivided into a number of different types, using different features to subclassify them. A sample of these systems in northern and central Vanuatu is given in (2) and in southern Vanuatu and New Caledonia in (3). As discussed in §1, innovative forms for the numerals are often compounds which include a form which is the same as, or is related to, the numerals 1–4. In (2) and (3) below, bolding marks the numerals 1–4 in the forms for 6–9.[6] Of the languages given here, Merei and Naman are imperfect decimal, Lewo and Sye are mixed, and the remainder are quinary.

(2)	**POc**	**Merei**	**Naman**	**SE Ambrym**	**Lewo**
		Santo	Malakula	Ambrym	Epi
1	various	*ese*	*savakh*	*tei*	*taaŋa*
2	*rua	*ruwa*	*iru*	*lu*	*lua*
3	*tolu	*tolu*	*itəl*	*tol*	*telu*
4	*pat(i)	*vat*	*i*	*ves*	*hat vari*
5	*lima	*lima*	*iləm*	*lim*	*lima*
6	*onom	*maravo*	*nsous*	***te**hesav*	*o**rai***
7	*pitu	*ravo**rua***	*nsu**ru***	***lu**hesa*	*o**lua***
8	*walu	*rap**tol***	*nsu**təl***	***ol**hesa*	*o**relu***
9	*siwa	*rait**at***	*nso**ves***	***hat**hesa*	*o**vari***
10	*sa-[ŋa]-puluq	*saŋavul*	*saŋavəl*	*he-xa-lu*	*lua-lima*
20	*rua-ŋa-puluq	*ŋavul-rua*	*(na)ŋavəl iru*	*hanu tap*	*lua-lima yam lua*

[5] Various forms have been reconstructed with the meaning 1 in POc, including *ta, *sa, *tai, *kai, and various combinations of these. In many languages, most numerals (though often not 1 and 10) have a historical or synchronic prefix, which often is, or is derived from, a 3SG or 3NSG verbal prefix: Uripiv in (1) illustrates this.

[6] Some relationships may not be immediately obvious: Lewo, for example, reflects POc *t as *t* initially (cf. 1, which is probably a compound, and 3) but as *r* medially (as in 6 and 8); the Southeast Ambrym form for 8 unexpectedly loses initial *t*; and so on.

The following comments can be made in relation to the data presented in (2) and (3). The form of the numerals 6–9 (or 7–9, since 6 is sometimes aberrant — see §3) may be:

- LIG-numeral (Merei, Naman, Lewo);
- numeral-LIG (Southeast Ambrym); or
- 5-(LIG)-numeral (Sye, Lenakel, Nemi, Xârâcùù).

The form of the numeral 10 may be:

- a synchronic monomorphemic form, generally a reflex of *sa-[ŋa]-puluq (Merei, Naman);
- a historical multiplicative, involving terms for 2 and 5 or 'hand' (Southeast Ambrym,[7] Lewo, Sye, Xârâcùù?);
- 5-(LIG)-5 (Lenakel); or
- an apparent compound, whose etymology is however unclear (Nemi?).

The form of the numeral 20 may be:

- a multiplicative involving terms for 2 and 10 (Merei, Naman, Lewo, Sye); or
- a phrase incorporating the word for 'man', 'person', or 'fingers/toes/digits' (Southeast Ambrym, Lenakel, Nemi, Xârâcùù). Lenakel *ieramim karena rəka*, for example, is literally man one he-is-not; Southeast Ambrym *hanu tap* is person whole; etc.

(3)	**POc**	**Sye**	**Lenakel**	**Nemi**	**Xârâcùù**
		Erromango	Tanna	N. New Cal.	S. New Cal.
1	various	*hai(teven)*	*karena*	*heec*	*chaa*
2	*rua	*nduru*	*kiu*	*heluk*	*baaru*
3	*tolu	*ndehel*	*kəsil*	*heyen*	*bachéé*
4	*pat(i)	*ndvat*	*kuvər*	*phoec*	*kêrêfùè*
5	*lima	*sukrim*	*katilum*	*nim*	*kêrênürü*
6	*onom	*mehikai*	*katilum-**karena***	*ni-bwe**ec***	*kêrênürü mê **chaa***
7	*pitu	*sukrim**nduru***	*katilum-**kiu***	*ni-bwe**luk***	*kêrênürü mê **baaru***
8	*walu	*sukrim**ndehel***	*katilum-**kəsil***	*ni-bwe**yen***	*kêrênürü mê **bachéé***
9	*siwa	*sukrime**ndvat***	*katilum-**kuvər***	*ni-bo**vac***	*kêrênürü mê **kêrêfùè***
10	*sa-[ŋa]-puluq	*narwolem*	*katilum-**katilum***	*paidu*	*duchêêxê*
20	*rua-ŋa-puluq	*narwolem nduru*	*ieramim karena rəka*	*hee kahok*	*xê chaa kamûrû*

2.2 Loyalty Islands

The Loyalty Islands languages follow a somewhat different pattern. The numerals 1–20 in the three Loyalties languages are given in (4).

These three languages show some similarities and some differences:[8]

---

[7] For Southeast Ambrym, Parker (1970) gives *hexalu* 10 (cf. *he-* 'hand', *lu* 2) and *hanu tap* (person whole) 20. Some modern speakers however have reanalysed this system, and give *hanutap tei* 10 and *hanutap lu* 20 The *he* in the ligature *-hesa(v)* in 6–9 is probably 'hand'.

- **Iaai** has two forms for 5 which appear to be in free variation. It has something which could roughly be translated as 'two hands', 'three hands' and 'four hands' for 10, 15 and 20 (*baa-* being a form which is used for naming fingers, though not the actual form for 'hand', which is *beñi-*). For 10 there is the alternative compound *li beñi-ta* (DUAL hand-1PL.INCL). Other numerals show the numerals 1–4 linked to *thabüŋ* 5, *li beñi-ta* 10 and *baa-kun* 15 by *ke nua* 'and again'.
- In **Drehu**, the numerals 5, 10 and 15 are formed on the base *-pi*, which may possibly be related to a form *pi* meaning 'last, after'. Although the meaning of *trii* in *trii-pi* 5 is unknown, 10 and 15 are 2-*pi* and 3-*pi*. Drehu is particularly interesting, not only in prefixing the numeral to the ligature, but also in having three separate ligatures: *-ŋömen ~ -ŋemen* for 6–9, *-ko* for 11–14, and *-qaihano* for 16–19. Parts of these ligatures can be identified: the *ŋe* of *-ŋömen ~ -ŋemen* is 'and', *ko* is 'towards', and the *qai* of *-qaihano* means 'coming from'. 20 is 'one person'.
- **Nengone** has 2 × 5 or 'two hands' for 10; 11–14 are additives to both of these (*ne* = 'and'). 15 = 'two hands and five again', 16–19 are additives to this (minus *yawe* 'again'), and 20 is 'one person'.

(4)	**Iaai**	**Drehu**	**Nengone**	
1	*xaca*	*caa(s)*	*sa*	
2	*lo*	*lue*	*rewe*	
3	*kun*	*köni*	*ten(e)*	
4	*væk*	*eke*	*ece*	
5	*baa-xaca, thabüŋ*	*trii-pi*	*sedoŋ*	
6	*thabüŋ ke nua* ***xaca***	***caa****-ŋömen*	*sedoŋ ne* ***sa***	
7	*thabüŋ ke nua* ***lo***	***lue****-ŋömen*	*sedoŋ ne* ***rewe***	
8	*thabüŋ ke nua* ***kun***	***köni****-ŋömen*	*sedoŋ ne* ***ten***	
9	*thabüŋ ke nua* ***væk***	***eke****-ŋömen*	*sedoŋ ne* ***ece***	
10	*baa-lo, li beñi-ta*	*lue-pi*	*rue sedoŋ*	*rue tubenin*
11	*li beñi-ta ke nua* ***xaca***	***caa****-ko*	*rue sedoŋ ne* ***sa***	*rue tubenin ne* ***sa***
12	*li beñi-ta ke nua* ***lo***	***lue****-ko*	*rue sedoŋ ne* ***rewe***	*rue tubenin ne* ***rewe***
13	*li beñi-ta ke nua* ***kun***	***köni****-ko*	*rue sedoŋ ne* ***ten***	*rue tubenin ne* ***ten***
14	*li beñi-ta ke nua* ***væk***	***eke****-ko*	*rue sedoŋ ne* ***ece***	*rue tubenin ne* ***ece***
15	*baa-kun*		*köni-pi*	*rue tubenin ne sedoŋ yawe*
16	*baa-kun ke nua* ***xaca***	***caa****-qaihano*	*rue tubenin ne sedoŋ ne* ***sa***	
17	*baa-kun ke nua* ***lo***	***lue****-qaihano*	*rue tubenin ne sedoŋ ne* ***rewe***	
18	*baa-kun ke nua* ***kun***	***köni****-qaihano*	*rue tubenin ne sedoŋ ne* ***ten***	
19	*baa-kun ke nua* ***væk***	***eke****-qaihano*	*rue tubenin ne sedoŋ ne* ***ece***	
20	*baa-væk*	*caatr*	*sa reŋom*	

[8] I am grateful to Claire Moyse for providing the Nengone data and assisting with much of the analysis of Drehu and Nengone.

2.3 Distribution of different types of innovative systems

The distribution of these various subtypes is given in Table 1.[9] Leaving aside some minor details for the moment, it seems that we can divide these languages into two major groups on the basis of the nature of the innovative system: a northern area comprising all of the Northern Vanuatu and some of the Central Vanuatu subgroups, where the systems are imperfect decimal, and a southern area comprising most of Southern Vanuatu and all of New Caledonian, where the systems are quinary. There is a third, intermediate group separated from the northern and southern groups by solid lines in the table, in which mixed systems occur. Shading in the northern and southern groups indicates features typical of each group. The symbol † beside the name of an island or island group indicates that conservative decimal systems are also found there.

Table 1: Distribution of different innovative systems

	NUMERALS 6–9			FORM FOR 10			FORM FOR 20	
	5-(LIG)-num	LIG-num	num-LIG	single morph	5-(LIG)-5	2 × 5/hand	2 × 10	person, digits
IMPERFECT DECIMAL								
Torres/Banks		✓		✓			✓?	
Maewo		✓		✓			✓?	
Pentecost †		✓		✓			✓?	
Santo†	few	most		✓			✓?	
Malakula†		✓		✓			✓?	
N and W Ambrym		✓		✓			✓?	
MIXED								
Epi-Efate		✓				✓	✓?	
Erromango	S	Ur, Ut				✓	✓	
QUINARY								
SE Ambrym/Paama		P	SEA			✓		✓
Tanna	✓				✓			✓
Aneityum		✓			✓			✓?
N. New Cal.	✓					✓?		✓
S. New Cal.	✓					✓?		✓
Loyalties	I, N		D			✓		D, N; I = 4×hand

The features of the two major areas are as follows. In the northern, imperfect decimal, area (which also houses some languages with decimal systems):

[9] Single-language abbreviations in Table 1 are: for SE Ambrym/Paama, P - Paamese and SEA - Southeast Ambrym; for Erromango, S - Sye, Ur - Ura and Ut - Utaha; and for the Loyalties, D - Drehu, I - Iaai and N - Nengone. Otherwise, ✓ indicates that this feature is found in all languages in the island(s) concerned.

- forms for 6–9 are almost universally LIG-numeral;
- the form for 10 is a single morpheme (synchronically), and almost everywhere derives from POc *sa-[ŋa]-puluq; and
- the form for 20 is a multiple of 2 and 10.

In the southern, quinary, area:

- forms for 6–9 are almost universally 5-LIG-numeral;
- the form for 10 is normally either 5-LIG-5, 2 × 5 or 2 × 'hand'; and
- the form for 20 is normally 'one person', 'person's digits', or some other phrase referring to a person, implying that the counting of ten fingers and ten toes is complete.

The intermediate area in the middle of Table 1, incorporating Epi, the Shepherds, Efate and Erromango, shows a mixed system and seems to be a transitional region between the northern and southern areas; it has no single morpheme for 10, but does show 'two tens/ hands' for 20.

2.4 Incipient innovative systems?

There is a group of contiguous languages — Araki in southern Santo and five or six languages in northern Malakula — which are clearly decimal, with reflexes of *onom 6, *pitu 7, *walu 8 and *siwa 9. In these languages, forms for 1–5 are either historically monomorphemic or else have a prefix which is or was the 3SG subject marker. However, the numerals 6–9 take a quite different prefix. I present data below from Araki and from five northern Malakula languages.

(5)	**POc**	**Araki**	**Nese**	**Botovro**	**Tirax**	**Malua Bay**	**Vovo**
1	various	*hese*	*sakhal*	*hual*	*hkhal*	*səkhal*	*hual*
2	*rua	*dua*	*rru*	*rue*	*ru*	*i/ru*	*rue*
3	*tolu	*rolu*	*til*	*til*	*til*	*i/til*	*til*
4	*pat(i)	*v'ari*	*v'at*	*v'at*	*vat*	*i/vat*	*vat*
5	*lima	*lim'a*	*line*	*lim'e*	*lin*	*i/ləm*	*lime*
6	*onom	*hai/ono*	*kh/on*	*h/on*	*khɔ/wɛn*	*kho/en*	*on*
7	*pitu	*hai/p'iru*	*kho/dit*	*huo/dit*	*khə/dit*	*i/bit*	*kho/dit*
8	*walu	*ha/ualu*	*kho/al*	*ho/al*	*khɛ/wɛl*	*o/wel*	*kho/al*
9	*siwa	*hai/sua*	*khe/sve*	*khe/sve*	*khɛ/siv*	*kha/səp*	*khe/hive*
10	*sa-[ŋa]-puluq	*saŋavulu*	*saŋav'il*	*haŋavul*	*hŋavil*	*saŋavil*	*haŋavil*

The forms for 6–9 in these languages contain a prefix which would derive from *kV-, whose possible origin will be discussed in more detail in §4.2.[10] In these languages, then, 6–9 are morphologically marked, in comparison with the unmarked 1–5. They thus resemble languages with imperfect decimal systems in treating numerals above 5 differently from those below it, even though the POc roots for 6–9 are retained. Whether this represents some kind of influence from neighbouring imperfect decimal languages, or the beginning of a change from decimal to imperfect decimal, is difficult to tell.

[10] Araki *h* is the regular reflex of *k (e.g. *huru* 'louse' < *kutu, *siho* 'kingfisher' < *siko).

2.5 Higher numerals

Higher numerals — by which I mean monomorphemic forms for 100, 1000 and the like — are found in a number of Oceanic languages with decimal systems. However, a number of languages in the northern part of Vanuatu with imperfect decimal systems also have forms for 100 and 1000 which are not compounds or paraphrases (though in the southern part of Vanuatu and in New Caledonia 100, for example, is usually 'five men'). For example:

(6)

	100	**1000**		**100**	**1000**
Mwotlap	*m̃eldēl*	*tey*	Uripiv	*nuŋut*	*evin, nuvesaŋ*
Mota	*m̃elnol*	*tar*	Neverver	*naŋat*	*netar*
Merei	*lavul*		Namakir	*na-ponati*	
Avava	*aŋat*	*atar*	Nguna	*p̃onotia*	*maanu*
Neve'ei	*naŋat*	*netar*	S Efate	*tifli*	*p̃on*
Naman	*noŋot*		Sye	*nalem*	
Tape		*itar*			

The Mwotlap and Mota forms for 100 are historically bimorphemic: François (2005:498) reconstructs an earlier *m̃ele-dolu 100 (lit. cycas palm-whole), with the second element probably deriving from PEOc *udolu 'very many, all' (Pawley 1972:55). The Malakula *ŋat-type forms for 100 may derive from Proto Malayo-Polynesian *sa-ŋa-Ratus 100 (Robert Blust, pers. comm. 4 Feb. 2009) with regular loss of *sa- and the final syllable (and fairly regular loss of *R): thus *sa-ŋa-Ratus > *-ŋat*. For 1000, PEOc *tari, which occurs in other languages meaning '(very) many', can be reconstructed (Pawley 1972:55; see also François 2005:500).

3 Forms for 'six'

Forms for 6 in innovative systems do not always follow the same pattern as forms for 7–9, and it will be worth discussing these briefly here, both in view of the discussion in §2 and also as an introduction to the discussion on ligatures in §4. Two examples of 'irregular' forms for 6 can be found in (2) and (3) above: Merei has *ma-ravo* (abbreviated form of 5 + LIG) for 6 but (underlying) *ravo* (LIG) + 2/3/4 for 7/8/9; and Sye has *mehikai* 6, but 7/8/9 are compounds of *sukrim* 5 + 2/3/4.[11]

A sample of 'irregular' forms for 6 is given in (7), and these are discussed in more detail below.

[11] I exclude from the 'irregular' category forms for 6 which basically follow the same pattern as 7–9 (in being composed of 5-1, 5-LIG-1 or LIG-1) but where the element representing 1 is a recognisably reduced or modified form of the numeral 1, as in Southeast Ambrym and Lewo in (2), or some other form meaning 'one' which is not used in serial counting and one or more others used within noun phrases.

(7)	**Wusi**	**Tolomako**	**Tasiriki**	**Tape**	**Avava**	**Ura**	**Anejom̃**
	Santo	Santo	Santo	Malakula	Malakula	Erromango	Aneityum
1	*ehe*	*tea*	*ʔese*	*isig*	*sapm*	*sai*	*ithii*
5	*lima*	*lina*	*lima*	*iləm*	*ilim*	*suworem*	*mijman*
6	*lima-rave*	*lina-rave*	*ha-ʔese*	*ləm-jis*	*sout*	*misai*	*meled*
7	*rave-rua*	*lina-rave-rua*	*ravaʔa-rua*	*ji-ru*	*sou-ru*	*sinelu*	*meled-erou*
8	*rap-tol*	*lina-rave-tolu*	*ravaʔa-tolu*	*ji-təl*	*se-tl*	*sinehli*	*meled-esej*
9	*ra-pati*	*lina-ra-tati*	*ravaʔa-vati*	*ji-vet*	*sa-vat*	*sinivat*	*meled-emanohowan*

Three different types of irregular forms for 6 are found in Santo, illustrated in (7) by Wusi, Tolomako and Tasiriki.

- The Wusi-type has 5-LIG for 6 but LIG-2/3/4 for 7–9. This type is found in most languages of Santo which have imperfect decimal systems, including various dialects of Central Santo and Southwest Santo, Kiai, Shark Bay, Mav'ea and Merei — i.e. much of the southern half of the island as well as the northeast.
- The Tolomako-type also has 5-LIG for 6, but 5-LIG-2/3/4 for 7–9. This type is also found in Sakao, which neighbours Tolomako in the Big Bay area, but in addition it occurs further to the south in South Central Santo.
- In the Tasiriki-type, 6–9 are all of the form LIG-1/2/3/4, but the ligatures are phonologically unrelated: Tasiriki *haʔ-*, Akei *awe-* with 6, but Tasiriki *ravaʔa-*, Akei *rava-* with 7–9. (Tasiriki and Akei are both dialects of Southwest Santo, though other dialects are of the Wusi-type).

Two different types of 'irregular' forms for 6 are found in Malakula, illustrated in (7) by Tape and Avava.

- Tape and V'ënen Taut (Big Nambas) have 5–1 for 6 but LIG-2/3/4 for 7–9.
- The Avava-type has the same LIG morpheme (given some vowel alternations) for all numerals 6–9. However, the form following LIG in 6 bears no phonological resemblance to the form for 1, nor to any other form with a similar meaning: Crowley (2006a:57) says of Avava *sout* 6 that it 'involves the same initial element *sV-* noted for the numerals 7–9, but with a following element that cannot be related to any of the other cardinal numerals'. Note Naman *nsous* in (2), Avava *sout* in (7) and the following in two other Malakula languages:

(8)		**Neve'ei**	**Neverver**		**Neve'ei**	**Neverver**
	1	*sevakh*	*skham*	6	*nsouh*	*jos*
	2	*iru*	*ru*	7	*nsu-ru*	*jo-ru*
	3	*itl*	*tl*	8	*nsu-tl*	*jo-tl*
	4	*ivah*	*vas*	9	*nsa-vah*	*jo-vas*
	5	*ilim*	*lim*	10	*naŋavil*	*naŋavul*

In Paamese, the form for 6 uses a different LIG from the one used with 7–9: note *lahi-tāi* LIG-1 'six', but *lau-lu, lau-tel* and *lau-hat* 7–9. Crowley (1982:98) suggests that the ligature for 6 is the verb *lahi* 'carry' and the other ligature is the noun *lau-* 'leaf'. I will return to this in the next section.

In Erromango, forms for 6 seem to derive from *ma 'and' + 1 (although Sye *mehikai* does not bear too close a resemblance to *hai(teven)* 1). Forms for 7–9 are based on *sukrim* 5 + 2–4 in Sye, but on LIG-2–4 in the other Erromangan language, as illustrated in (9).[12]

(9)

	Ura	**Utaha**		**Ura**	**Utaha**
1	*sai*	*soɣoi*	6	*misai*	*miseɣai*
2	*gelu*	*kalu*	7	*sine-lu*	*simna-lu*
3	*gehli*	*kihili*	8	*sine-hli*	*simni-heli*
4	*lemelu*	*lemelu*	9	*sini-vat*	*simni-vat*
5	*suworem*	*sukrim*	10	*lurem, durem*	*narolem*

The Anejom̃ form for 6 is just LIG: *meled* 6 is composed of *m-* 'echo subject prefix' + *eled* 'be left over'. Numerals 7–9 are composed of *meled*-2/3/4.[13]

Forms for 6 thus often pattern differently from forms for 7–9. This is an important element in the hypothesis of the development of the numeral systems which I will develop in §4 and §5.

4 Ligatures

The ligature used in those quinary languages in the southern area which do not simply conjoin 5 and another numeral is either a reflex of POc *ma 'and', some other coordinating conjunction, or a phrase like 'and more' or 'and again'. This is not of any particular interest. Of more interest historically is the type of ligature used in the northern languages, where the numeral 5 normally does not appear (except in 6 and in the Tolomako-type described above), and where the ligature is *not* a coordinating conjunction. Three reasonably widespread ligatures can be identified.

4.1 PNCV *lave-a

The most widespread ligature in Vanuatu has the form *lavV-* or *ravV-*. Cognates can be seen in Merei in (2) and in Wusi, Tolomako and Tasiriki in (7). The distribution of related forms is as follows:

- all Torres, Banks and Maewo languages;
- all Pentecost languages (except for the decimal Raga);
- all Santo languages which have imperfect decimal systems;

[12] Ura and Utaha have lost *pat(i) for the numeral 4 and replaced it with an additive form 2-and-2. However, a reflex of *pat(i) shows up in the compound form 9.

[13] Anejom̃ *d* is a voiceless interdental fricative /θ/, and *eled* appears to be related to PNCV *malazi 'leftover food' (with probably a stative prefix *ma-).

- some Malakula languages which have imperfect decimal systems: Axamb, Lendamboi, Maskelynes, Banam Bay, Unua and Aveteian. These are spoken mainly in south and east Malakula, though there is no particular genetic connection between just these languages, as they belong to two different first-order subgroups of Malakula languages (Lynch 2007), and some of their close relatives behave differently; and
- Paamese.

In addition, the following are probably also related:

- the type *milip-* or *miliv-* in a number of Ambrym languages, which may represent the *lavV-* form preceded by a subject-TAM marker, or by a conjunction, or possibly by a reduced form of *lima 5; and
- the form *la-* in the languages of the Shepherds and Efate, in which the second syllable was lost.

Some additional examples of the forms for 6–9 in one language from each of these areas follow:

(10)	**Mwotlap**	**Peterara**	**Sowa**	**Kiai**	**Lendamboi**	**Fali**	**Nguna**
	Banks	Maewo	Pentecost	Santo	Malakula	Ambrym	Shepherds
6	*leve-te*	*lav-tuɣ^{w}ale*	*lo-wal*	*lima-rave*	*i-lav-sua*	*milip-tʃe*	*la-tesa*
7	*livi-yō*	*lav-rua*	*lew-ru*	*rav-ua*	*i-lav-rua*	*milip-ru*	*la-rua*
8	*leve-tēl*	*lav-tolu*	*lep-tul*	*ra-tolu*	*i-lavu-təl*	*milip-tʃil*	*la-tolu*
9	*leve-vet*	*lei-vati*	*lak-pat*	*ra-pati*	*i-la-vas*	*milip-fer*	*lo-veti*

There is a certain amount of erosion of the second syllable in a number of areas, and also a certain amount of variation in the vowels, much of it possibly due to some kind of assimilation to the vowel of the numeral root. Crowley's comment in §3 above regarding Paamese is a case in point: though he suggests that *lahi-* used with 6 and *lau-* with 7–9 have different meanings/origins, I suggest that they both derive from the same form, with loss of *v and a vowel change in the latter. The examples in (10) also illustrate these kinds of changes: Mwotlap shows vowel alternations, Peterara has loss of *v in 9, Sowa has vowel and consonant changes, Lendamboi loses *v in 9, and Nguna shows loss of the second syllable plus vowel changes.

The initial consonant reflects *l in most of the languages in which this occurs, although some Santo languages suggest initial *r rather than *l. Phonologically conservative languages (conservative in the sense that they do not show significant vowel changes) suggest that the first vowel was *a; and it is clear that the second consonant was PNCV *v, POc *p.

There is more of a problem deciding what the second vowel was, since many languages have lost it in some or all numerals or show vocalic changes. An initial approach to this problem was to examine only those languages in which the form remained constant from 6–9 (or 7–9 if 6 is aberrant). These languages are listed in regular font in Table 2. Languages in italics in that table are those which have a regular prefix-final vowel in all numerals except 9, where dissimilation occurs because of two occurrences of *v in adjacent syllables (e.g. Marino underlying /leva-vati/ > *la-vati*).

Table 2: Final vowels in the ligature *lavV- / ravV-*

VOWEL	LANGUAGES
-i	Banks: Vera'a
-ia	Banks: *Merlav*
-e	Banks: Lehali, Lehalurup, Nume, Dorig. Santo: *Tolomako, Navut, Shark Bay*
-ea	Banks: Mota
-a	Santo: Akei, *Marino*
-aʔa	Santo: Tasiriki
-u	Santo: *Malmariv*

François (2005) has carried out a detailed study of the reflexes of the POc vowels in Banks languages. Most of these languages have undergone significant sound changes, with only Mota being conservative. The sources of the prefix-final vowel in some of the Banks languages (cf. François 2005:490–491) are possibly as follows:

(11) Lehali *e*: *i(C)e, *i(C)a, *i(C)o, *e(C)i, *e(C)u
Lehalurup *e*: *e(C)i, *e(C)u
Nume *e*: *e(C)e, *e(C)a, *e(C)o
Dorig *e*: *e(c)e, *e(C)a, *e(C)o
Mota *ea*: *e(C)a

This suggests to me that the original form may have been *lave(C)-a,[14] i.e. a transitive verb with an object suffix, which correlates with Crowley's suggestion regarding the Paamese ligature *lahi-* deriving from 'carry'.

Now Clark (in press) has reconstructed two PNCV verbs which may be variants of the same original form: *lavi 'carry, take' and *la-i 'take, give'. It seems possible that this linker *lave-a is related to, or is an aberrant form of, *lavi-a 'carry/take-3SG'. I assume that the Tolomako-type discussed in §3 was probably the original: that is, the forms for 7–9 were *lima-lave-a-rua/tolu/vat(i), while the form for 6 was either simply *lima-lave-a or *lima-lave-a-ta (or some other form for 1). The meaning would have been something like '5 carry 1/2/3/4'. In most languages with this ligature, there was a certain amount of redundancy in the trimorphemic form, and this generally reduced to a bimorphemic form, with the numeral 1 being dropped from 6 and *lima 5 being dropped from 7–9.

An apparent reflex is found outside Vanuatu in at least one language: Tolai (Kuanua) of East New Britain has *lap-tikai, lavu-rua, lavu-tul* and *lavu-wat* for the numerals 6–9 (compare *tikai, au-rua, au-tul, ai-wat* for 1–4). Lynch, Ross and Crowley (2002:72) proposed POc *(l,r)apo- to account for this correspondence, but it is likely that this resemblance is accidental, and that the Tolai prefix may derive from *lap ~ lav* 'follow'.

4.2 *[la]kau-

There is a number of languages which have a ligature which is, or contains, *kV-. Some of these were listed in (5) in the discussion in §2.4 of 'incipient innovative systems' in south Santo/north Malakula. An apparently related form is the ligature *ha-* in Tasiriki *ha-*

[14] Pawley (1972:47) reconstructed this form as *lapu-.

ʔese 6 (see §3 above).[15] Seemingly similar forms are found in southeast Malakula (Aulua, Port Sandwich and Avok below) and Epi (Lewo in (2), where *k > Ø, and Bonkovia and Mae-Morae below), most with a preceding syllable.

(12)	**Aulua**	**Pt. Sandwich**	**Avok**	**Bonkovia**	**Mae-Morae**
6	*dro-vokhol*	*emut-sukai*	*ŋə-mekh-cəkai*	*wo-ra*	*lok-rogai*
7	*drokhu-rue*	*e-mokh-ü*	*ŋə-mekhu-ru*	*oko-lua*	*loku-lua*
8	*drokh-til*	*e-mokhu-röj*	*ŋə-mekho-rœr*	*oko-rolu*	*lok-rol*
9	*drokh-ves*	*e-mokhu-pac*	*mekho-pæc*	*oko-veri*	*lak-var*

These forms suggest *ko- or *ku-, with a preceding element. In Aulua and a number of Epi languages (like Mae-Morae in (12)), the preceding element is a liquid + vowel, and I suggest here that the original form of this ligature may be the PNCV verb *lakau (~*lakawa) 'cross over' (Clark 2009). This makes sense semantically: having counted the fingers of one hand, one then 'crosses over' to the other hand to begin counting at 6. This is semantically (though not phonologically) identical to the ligature *bena* in Banoni and Piva in Bougainville (Lincoln in press): *bena* 'cross over' is 6, and *bena*-2/3 are 7–8 (9 is *visa*, possibly a metathesised form of *siwa).

The ligature *lakau- presumably reduced to *kau- in a number of languages. Note its phonological similarity to the forms in the 'incipient innovative' languages given in §2.4 (Araki *hai-*, Nese *kho-, khe-*, Botovro *ho-, khe-*, Tirax *khɔ-, khɛ-*, Malua Bay *kha-, kho-* and Vovo *khe-, kho-*).

4.3 *zau-

There is a group of languages in Malakula which have a ligature whose form was possibly *zau-, with initial PNCV *z < POc *j, (although the vocalic element undergoes some variation in some of these languages). Some examples:[16]

(13)	**Avava**	**Neve'ei**	**Naman**	**Neverver**	**V'ënen Taut**	**Tape**	**Nāti**	**Nahavaq**
6	*sout*	*nsouh*	*nsous*	*jos*	—	—	*seu-siʔ*	*sow-siʔ*
7 /_*rua	*sou-*	*nsu-*	*nsu-*	*jo-*	*sa-*	*ji-*	*seu-*	*sow-*
8 /_*tolu	*se-*	*nsu-*	*nsu-*	*jo-*	*sa-*	*ji-*	*seu-*	*sow-*
9 /_*pat(i)	*sa-*	*nsa-*	*nso-*	*jo-*	*sa-*	*ji-*	*seu-*	*sow-*

A possible origin for this form — though I am not at all confident about this — may be PNCV *sabo. Clark (in press) glosses this as 'ignorant, incompetent, lost', though in some languages the meaning has shifted to 'other, different', as in Paamese *savo* 'different', *mee-savo* 'someone else, stranger', Naman *i/nsəb* 'other, different'; so 7 would be 'other (hand)-2'. It is also not clear whether this form is related in any way to the ligature found in two Erromangan languages given in (9): Ura *sine-, sini-* and Utaha *simna-, simni-* (and note Ura *sai* 'other, different').

---

[15] The regular reflex of *k in Tasiriki is *ʔ*, but *kVk seems to be reflected as *hVʔ*; cf. 'Proto Santo' *ka-kara 'red' (Wusi *kara*, Nokuku *kekara*, Tolomako *ɣaɣara*) > Tasiriki *haʔara*.

[16] V'ënen Taut and Tape do not use this form in the numeral 6, which is 5–1. In Nāti and Nahavaq, the form for 6 is transparently LIG-1, but this is not the case in the other languages, and the form for 6 is given in full in the first four languages in (13).

5 Discussion

Two different general types of innovative systems seem to have developed in the Vanuatu–New Caledonia area — imperfect decimal and true quinary systems — with a mixed system occurring in the central area. The basic forms of relevant numerals in these systems are given in Table 3.

Table 3: General types of innovative system**s**

	Imperfect decimal			**Mixed**	**Quinary**
	*lave-a-	*[la]kau-	*zau-		
6	5-carry(-1)	cross.over-1	other(-1)	*same*	5(-and)-1
7	(5-)carry-2	cross.over-2	other-2	*as*	5(-and)-2
8	(5-)carry-3	cross.over-3	other-3	*imperfect*	5(-and)-3
9	(5-)carry-4	cross.over-4	other-4	*decimal*	5(-and)-4
10	ten			two fives/hands	5(-and)-5, two fives/hands
20	two tens			two tens	one person

5.1 Imperfect decimal, quinary and mixed systems

The imperfect decimal type is found throughout north and central Vanuatu (except for those languages with pure decimal systems) as far south as the middle of Ambrym. It is in fact basically a decimal system, with its only deviation being that it has replaced the terms for 6–9 with compounds, but compounds of a quite different type from those found in the far south. Many of these languages also have uncompounded forms for 100 and 1000. (Indeed, there are even relics of original terms for 6–9 in some of these languages. In an intriguing footnote to his discussion of the quinary Lewo numeral system, illustrated in (2) above, Early (1994:213) says:

> The Lewo counting system for spirit-creatures is still widely known, retained as an item of conscious cultural knowledge, but without any current function. The numbers, which go from 1 to 10, are: *taka, luaka, telka, verka, limka, kona, isi, varo, siwe, kuru*.

The numerals 6–9 reflect POc *onom, *pitu, *walu and *siwa.)

True quinary systems are found to the south of this area. In these systems, numerals effectively stop at five, and everything above that is an additive compound until twenty is reached, which is usually a compound meaning ‘one person’ — i.e. all fingers and toes have been counted. Addition then starts again from 21–40 (‘two people’), 41–60 (‘three people’), etc. There are no monomorphemic words for 100 or 1000. Mixed systems occur in parts of central Vanuatu (Epi, Shepherds and Efate and Erromango).

Given that true or complete decimal systems still survive in parts of northern Vanuatu, and given the distribution of the imperfect decimal and quinary types within Vanuatu and New Caledonia, the following sequence of events suggests itself to me:

I. The original settlers of the area (somewhere in northern Vanuatu) had a true decimal system.

II. The first innovation was to change to an imperfect decimal system, using some variant of LIG-1–4 for 6–9, but otherwise leaving the decimal system unchanged.

III. A true quinary system developed out of this, somewhere in the centre or south of Vanuatu, and the two systems were in some kind of competition for some time, which resulted in the mixed systems in the central area.

Proposal I is uncontroversial: the full decimal system is retained in at least some Vanuatu languages, which are geographically non-contiguous (including the Lewo counting system for spirit-creatures).

Proposal II is also uncontroversial. Settlement of Vanuatu was generally in a north–to–south direction. The imperfect decimal system is widespread in the north, and elements of it show up as far south as Anejom̃ in the extreme south. The use of the LIG-1–4 system probably evolved from counting on the fingers of one hand and then crossing over to the other (cf. also Lincoln in press); the bare use of the ligature here (rather than 5-LIG-1–4), along with the fact that the numeral 6 is often different in form from 7–9 and indicates a 'crossing over', tend to support this. The fact that some of these languages preserve the ligature *dum^{w}a- for 11–19, found also in languages with true decimal systems, along with words for 100 and 1000 (which may, of course, have come to mean 'fairly high number' and 'somewhat higher number still'), lends support to this hypothesis.

Proposal III recognises that quinary systems have a restricted geographical distribution, with no evidence of any being found anywhere north of Southeast Ambrym.

5.2 On the origins of the systems

As I mentioned at the beginning of this paper, Blust (2005) has suggested that the 'quinary' systems — by which I assume he means innovative (in my terminology), not just quinary — in this area arose out of Papuan contact *in situ*. That suggestion, of course, is impossible to disprove, since it implies also that any Papuan language or languages spoken in Vanuatu–New Caledonia which influenced the Oceanic languages of this region have since disappeared.

The nearest extant Papuan languages are in the central Solomon Islands. All Oceanic languages of the Solomon Islands (excluding Bougainville) as far south as Makira have pure decimal systems. Michael Dunn (pers. comm. 3 October 2007) says of the Solomons Papuan languages that:

> there aren't any quinary systems there any more. All the Solomons Papuan languages have productive decimal systems, but there's internal evidence … that they developed out of a quinary system.

There are, however, quinary systems close to northern Vanuatu, in the languages of the Temotu Province of Solomon Islands, all of which are now known to be Oceanic and not Papuan (Lincoln 1978, Ross and Næss 2007). There are nine languages in this area: all three in Vanikoro and one in Utupua have straightforward decimal systems. Of the remainder, Nagu in Santa Cruz has a subtractive system (7 = LIG-3, 8 = LIG-2, 9 = LIG-1); and Äiwoo in the Reef Islands, Natügu in Santa Cruz and two languages in Utupua have innovative systems of the same general type as northern Vanuatu (i.e. 7 = LIG-2), though Natügu in (14) appears to have 2-LIG. The relevant numerals in these languages are listed in (14) (1–9 from Ross and Næss 2007, 10 from Tryon and Hackman 1983).

These languages also appear to have forms for 100 and 1000 which are not compounds (Tryon and Hackman 1983). In other words, with the exception of the subtractive Nagu, these are of a similar type to the imperfect decimal systems in much of Vanuatu, though there do not appear to be any cognates among the ligatures.

Lean (1992) conducted a thorough study of Papuan (and Oceanic) counting systems on the island of New Guinea and neighbouring islands. His conclusions as to the distribution of various kinds of systems are set out in Table 4 (Table 63 in the original). While recent research has suggested that the East Papuan phylum is no longer a valid grouping (Ross 2001), I leave that heading in the table (with that column shaded) since it encompasses the extant Papuan languages geographically closest to Vanuatu and New Caledonia. The table also conflates some types which differ in only very minor ways, and uses different names for some of the types of systems to conform with the usage in this paper. The two innovative systems found in Vanuatu–New Caledonia are also shaded on this table.

(14)	**Äiwoo**	**Natügu**	**Nagu**	**Nebao**	**Tanibili**
	Reef Is.	Santa Cruz	Santa Cruz	Utupua	Utupua
1	*nyigi*	*tʌe-sʌ*	*tëte/ëte*	*tua*	*suo*
2	*li-lu*	*li*	*la-li*	*l-lu*	*bu-yu*
3	*eve*	*tü*	*lʌ-tü*	*t^hɔ*	*bo-k^wo*
4	*u-vä*	*p^wä*	*lʌ-fɔ*	*hia*	*mā-piɔ*
5	*vi-li*	*nʌlvü*	*lʌ-mëf*	*haŋi*	*kavili*
6	*pole-gi*	*e-sʌ-mʌ*	*lʌ-mët^{h}emë*	*uru*	*kavili suɔ*
7	*pole-lu*	*ë-li-mʌ*	*tumë-tu*	*va-lu*	*suo-vi-yo*
8	*pole-e*	*ë-tü-mʌ*	*tumë-li*	*va-ro*	*ve-vi-ro*
9	*polo-uvä*	*ë-p^wä-mʌ*	*tumë-te*	*wa-hia*	*vere-ve-pio*
10	*nugolu*	*nʌpnu*	*nëpnu*	*ŋalɔ*	*vere-ŋalu*

Table 4: Distribution of Papuan numeral systems

Types	West Papuan	Torricelli	Sepik-Ramu	Trans New Guinea	East Papuan	Minor Phyla	**Total**
Binary	0	0	3	39	0	0	**42**
Binary + 5	0	24	13	134	2	3	**176**
Quinary	0	2	17	52	1	7	**79**
Base–4	0	0	1	6	0	2	**9**
Base–6	0	0	0	5	0	0	**5**
Body parts	0	0	8	58	0	4?	**70?**
Imperfect decimal	7	0	3	8	12	3	**35**
Decimal	3	0	1	3	8	0	**16**

Lean (1992, §3.4) notes that the imperfect decimal systems of the 'East Papuan phylum' are mainly located in New Britain and Bougainville. But Lincoln (in press) has clearly demonstrated that the imperfect decimal systems of Banoni and Piva (and, probably, other

Bougainville Oceanic languages) have *not* developed as a result of Papuan influence, but as a result of the physical nature of the counting process. I believe I have demonstrated that the same is true of the imperfect decimal systems in Vanuatu–New Caledonia. The only point at issue, then, is how the quinary systems developed; these differ from imperfect decimal systems mainly in the loss of forms for 10 and in using the person-type construction rather than numerals for 20. Was this due to contact? If not, and this seems highly unlikely given their location, was it simply due to the 'regularisation' of LIG-1–4 systems as 5 + 1–4 systems? Proposal III above also referred to the two systems being in 'competition' for some time. Just as Araki and the Malakula languages discussed in §2.4 seem to be showing a change from decimal to imperfect decimal, so too the data in (14) suggest a change in progress from imperfect decimal to quinary; but in both cases the change is incomplete. The ligatures (*lave-a and the others) lost their transparency as 'carrying' or 'cross over' verbs, partly due to phonological change and attrition, and were replaced by a more transparent compound 5 + 1–4.

6 Conclusion

In this paper, I have described in some detail the numeral systems of the languages of Vanuatu and New Caledonia, and shown how they developed morphologically out of the POc decimal system which is still found in some northern Vanuatu languages. I have not attempted to account for the origin of these systems, or for the motivation which led to a change from decimal to innovative systems, but have provided adequate data so that the Papuan contact hypothesis might be tested. When it is tested, however, it is worth noting that the greatest deviation from the standard Oceanic decimal system comes in those languages which are geographically most distant from known Papuan-speaking areas.

Appendix: Orthography and sources of data

Standard orthography is used in most cases (but not for languages which do not have one), though the velar nasal and glottal stop are consistently represented as *ŋ* and *ʔ* and the schwa in Vanuatu languages as *ə*. The symbols *m'*, *v'* and *p'* in various languages are apicolabials; labiovelars (often marked by tildes in some Vanuatu languages) are consistently written with a superscript *w*; *kh* in Malakula languages represents a velar fricative (variably voiced depending on position), irrespective of what the standard orthography uses; Araki *d* is a flap, *r* is a trill, whereas in Nese and Uripiv *r* is a flap and *rr* a trill.

Specific data sources for many of the languages discussed in this paper are given below where there is no specific reference in the text. Data for those which do not appear on this list, and additional data for some which do, come from either Tryon (1976), Charpentier (1982), Lynch, Ross and Crowley (2002), Clark (in press) and/or the *Austronesian Basic Vocabulary Database* (Greenhill, Blust and Gray 2005–09). Data for Nahavaq, Neverver, Tirax and Uripiv come from unpublished wordlists/dictionaries compiled by Laura Dimock, Julie Barbour, Amanda Brotchie and Ross McKerras respectively, while Nengone data are from Claire Moyse.

Sources for other languages are: Anejom̃ (Lynch 2000); Araki (François 2002); Avava (Crowley 2006a); Drehu (Moyse-Faurie 1983); Iaai (Ozanne-Rivierre 1976); Lenakel (Lynch 1978); Lewo (Early 1994); Merei (Chung 2005); Mota (Codrington 1885);

Mwotlap (François in prep.); Naman (Crowley 2006b); Nāti (Crowley 1998b); Nemi (Haudricourt and Ozanne-Rivierre 1982); Nese (Crowley 2006c); Paamese (Crowley 1982); Port Sandwich (Charpentier 1979); South Efate (Thieberger 2006); Southeast Ambrym (Parker 1970); Sye (Crowley 1998a); Tape (Crowley 2006d); Ura (Crowley 1999); Utaha (Lynch 2001); V'ënen Taut (Fox 1979); Xârâcùù (Moyse-Faurie 1995).

References

Blust, Robert. 2005. Review of John Lynch, Malcolm Ross and Terry Crowley, *The Oceanic languages*. Richmond, UK: Curzon Press, 2002. *Oceanic Linguistics* 44.544–558.

Charpentier, Jean-Michel. 1979. *La langue de Port-Sandwich (Nouvelles-Hébrides): Introduction phonologique et grammaire*. Paris: SELAF.

—— 1982. *Atlas linguistique du Sud-Malakula/Linguistic atlas of South Malakula (Vanuatu)*. Paris: CNRS/L'Agence de Coopération Culturelle et Technique. [Langues et cultures du Pacifique 2.]

Chung, Yun Shing Anthony. 2005. *A descriptive grammar of Merei (Vanuatu)*. Canberra: Pacific Linguistics.

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.

—— 2009. **Leo Tuai: a comparative lexical study of North and Central Vanuatu languages.* Canberra: Pacific Linguistics.

Codrington, R.H. 1885. *The Melanesian languages*. Oxford: Clarendon Press.

Comrie, Bernard. 2008. Numeral bases. In Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie, eds *The world atlas of language structures online*, cht. 131. Munich: Max Planck Digital Library. http://wals.info/feature/131.

Crowley, Terry. 1982. *A dictionary of Paamese*. Canberra: Pacific Linguistics.

—— 1998a. *An Erromangan (Sye) grammar*. Oceanic Linguistics Special Publication No.27. Honolulu: University of Hawai'i Press.

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

—— 1999. *Ura: a disappearing language of southern Vanuatu.* Canberra: Pacific Linguistics.

—— 2006a. *The Avava language of central Malakula (Vanuatu)* [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.

Early, Robert. 1994. A grammar of Lewo, Vanuatu. PhD thesis. Canberra: The Australian National University.

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

François, Alexandre. 2002. *Araki: a disappearing language of Vanuatu*. Canberra: Pacific Linguistics

—— 2005. Unraveling the history of the vowels of seventeen Northern Vanuatu languages. *Oceanic Linguistics* 44.443–504.

—— In prep. Dictionnaire mwotlap–français. Electronic manuscript. Paris: LaCiTO, CNRS.

Greenhill, Simon, Robert Blust and Russell Gray. 2005–09. *The Austronesian Basic Vocabulary Database.* http://language.psy.auckland.ac.nz/austronesian.

Haudricourt, André-Georges and Françoise Ozanne-Rivierre. 1982. *Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie): Pije, fwâi, nemi, jawe.* Paris: SELAF.

Lean, Glendon Angove. 1992. Counting systems of Papua New Guinea and Oceania. PhD thesis. Lae: Papua New Guinea University of Technology. http://www.uog.ac.pg/glec/thesis/thesis.htm

Lincoln, Peter C. 1978. Reef-Santa Cruz as Austronesian. In Lois Carrington and S.A. Wurm, eds *Second International Conference on Austronesian Linguistics: Proceedings*, 929–67. Canberra: Pacific Linguistics.

—— in press. Count not on substrata. In John Bowden, Nikolaus Himmelmann and Malcolm Ross, eds *A journey through Austronesian and Papuan linguistic and cultural space: papers in honour of Andrew K. Pawley*. Canberra: Pacific Linguistics.

Lynch, John. 1978. *A grammar of Lenakel*. Canberra: Pacific Linguistics.

—— 2000. *A grammar of Anejom̃*. Canberra: Pacific Linguistics.

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

—— 2007. The Malakula linkage of central Vanuatu. Paper presented to the Seventh International Conference on Oceanic Linguistics, Nouméa, July 2–6.

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

Moyse-Faurie, Claire. 1983. *Le drehu, langue de Lifou (Iles Loyauté).* Paris: SELAF.

—— 1995. *Le xârâcùù, langue de Thio-Canala (Nouvelle-Calédonie): Éléments de syntaxe.* Paris: Peeters.

Ozanne-Rivierre, Françoise. 1976. *Le iaai, langue mélanésienne d'Ouvéa (Nouvelle-Calédonie): phonologie, morphologie, esquisse syntaxique.* Paris: SELAF.

Parker, Gary J. 1970. *Southeast Ambrym dictionary*. Canberra: Pacific Linguistics.

Pawley, Andrew. 1972. On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol. 3, 1–142. Honolulu: Bernice Pauahi Bishop Museum.

Ross, Malcolm. 2001. Is there an East Papuan phylum? Evidence from pronouns. In Andrew Pawley, Malcolm Ross and Darrell Tryon, eds *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton*, 301–321. Canberra: Pacific Linguistics.

Ross, Malcolm and Åshild Næss. 2007. An Oceanic origin for Äiwoo, the language of the Reef Islands? *Oceanic Linguistics* 46(2).456–498.

Thieberger, Nicholas. 2006. *A grammar of South Efate: an Oceanic language of Vanuatu*. Oceanic Linguistics Special Publication No.33. Honolulu: University of Hawai'i Press.

Tryon, Darrell. 1973. Linguistic subgrouping in the New Hebrides: a preliminary approach. *Oceanic Linguistics* 12.303–351.

—— 1976. *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.

Tryon, D.T. and B.D. Hackman. 1983. *Solomon Islands languages: an internal classification*. Canberra: Pacific Linguistics.

21 *Spread of the Saliba-Logea plural marker*

ANNA MARGETTS

1 Introduction

In Oceanic languages, number is generally not an inflectional category of nouns and an unmarked noun can have singular or non-singular reference. For Proto Oceanic, Lynch, Ross and Crowley (2002:70) reconstruct a system where '[o]nly some human nouns were (optionally) marked for number, probably by reduplication'. In modern Oceanic languages there is commonly a hierarchy where human or animate nouns are more likely to allow or require number marking than inanimate nouns. Often it is only a subclass of human nouns, such as kinship terms, which takes number marking and commonly this marking is not obligatory. The types of strategies found to express plural marking on nouns include reduplication, root modification with lengthening of vowels, and, more rarely, affixation (Lynch et al. 2002:37–39).

This paper details the number marking strategies of Saliba-Logea noun phrases which involves suffixation and, for a small number of nouns, also stem reduplication. The focus of the paper is on grammatical changes affecting the selection restrictions of the plural suffix and its spread into novel contexts.[1]

2 Plural marking in Saliba-Logea

Saliba-Logea is a Western Oceanic language of the Papuan Tip cluster spoken in Milne Bay Province of Papua New Guinea. It belongs to the Suauic linkage of the Nuclear Papuan Tip linkage (Ross 1988:190; Lynch et al. 2002:881–882). The two main dialects, Saliba and Logea, differ lexically but not grammatically. Mosel (1994:10) describes the plural suffix in Saliba as marking nouns referring to human beings and reports that plurality is for some human nouns marked by both the plural suffix and reduplication.

[1] I would like to thank all the speakers of Saliba and Logea who have contributed to the corpus used in this study and who helped with transcription and analysis. The research which lead to the creation of the text corpus was made possible by a grant from the DOBES program of the VW Foundation, Germany. I am grateful to Andrew Margetts for his help in manipulating the Toolbox concordance to make more complex searches possible. I am also grateful to Beth Evans and an anonymous referee for very helpful comments on an earlier version of this paper.

Margetts (1999:20–21) states that only humans and 'higher animals' (cf. Lichtenberk 1982) are marked for number and that for this class the singular form is unmarked but the plural form is suffixed.

In other languages of the Nuclear Papuan Tip linkage, plural marking seems to be restricted to kinship terms and involves either reduplication and/or a suffix which is cognate with the Saliba-Logea marker.

For (Kwato) Suau, Abel (n.d.:2) notes that 'singular and plural are usually the same except for the case of relationships, when *eao* is added, and sometimes the first syllable is doubled.' He provides the example *natu-gu* 'my child' versus *na-natu-gu-eao* 'my children'.

In Tawala, only the class of kinship terms marks plural productively and it does so by reduplication (Ezard 1997:54).

Dobu has a suffix with the allomorphs *-ao, -o*, and *-yao*, which marks plural on nouns. Lithgow (1975:32) describes that it is not obligatory and tends to be omitted when the plural meaning is obvious from context. Lithgow does not state whether the suffix is restricted to a particular class of nouns but the examples he provides are both kinship terms: *tubu-na-o* 'her grandparents' and *tasi-na-o* 'his brothers'. He also notes that the suffix is probably derived from the word *yau* 'many'.

In Saliba-Logea there is a plural suffix *-ao* with the allomorphs *-o, -wao, -yao*.[2] The allomorphs are to some extent predictable by phonotactics, but not entirely so. The allomorph *-o* occurs on nouns or suffixes ending in *-a*. The allomorph *-yao* occurs after the possessive suffix *-di*, and *-wao* after nouns ending in *-u.* But, nouns ending in *-i* and *-e* take the allomorph *-o* rather than *-yao*, as would be expected based on phonotactics. Sometimes speakers use *-yao* where *-wao* would be expected phonotactically and vice versa. Consider the examples in (1) and (2) which show the allomorphs *-yao* and *-wao* respectively following the possessive suffix *-mai* (first person plural exclusive).[3] Also note that the Suau form *-eao* [-yao] is not described to have allomorphs and according to Abel (quoted above) attached to *u*-final words where *-wao* may be expected.

(1) *sina-mai-yao*
mother-1EX.POSS-PL
our mothers

(2) *maiya-mai-wao*
with.3SG-1EX.POSS-PL
(s)he with us

As Mosel (1994) and Margetts (1999) point out, in Saliba-Logea plural marking of human nouns is obligatory. (In addition, domesticated or anthropomorphised animals can at times be marked with the suffix.) Consider the examples in (3) and (4).

(3) *wawaya*
child
'child'

wawaya-o
child-PL
'children'

(4) *ka-gu* *kaha*
CLF-1SG.POSS sibling
'my sister'

ka-gu *kaha-o*
CLF-1SG.POSS sibling-PL
'my sisters'

---

[2] Evidence that the plural marker is a suffix rather than a clitic includes the fact that it generally causes a stress shift, and that nothing can intervene between the noun and the plural marker apart from bound morphology.

[3] In addition to the Leipzig Glossing Rules the following abbreviations are used: ASSOC - associative; CONJ - conjunction; PP - postposition; RED - reduplication.

For some nouns, plural marking involves reduplication of the first two stem syllables, but in these cases the plural suffix is also required. Nouns with this pattern include *sine* 'woman', *tau* 'man',[4] *sinebada* 'old woman', *taubada* 'old man', *hewali* 'young man', *hasala* 'young woman' and *tanowaga* 'boss, owner'.

(5)	*sine*	*sine-sine-o*
	woman	RED-woman-PL
	'woman'	'women'
(6)	*taubada*	*tau-taubada-o*
	old.man	RED-old.man-PL
	'old man'	'old men'
(7)	*hewali*	*hewa-hewali-o*
	young.man	RED-young.man-PL
	'young man'	'young men'

If a noun with a human referent is directly possessed, as in the case of many kinship terms, the plural suffix follows the possessive suffix:

(8)	*natu-gu*	*natu-gu-wao*
	child-1SG.POSS	child-1SG.POSS-PL
	'my child'	'my children'
(9)	*yaya-da*	*yaya-da-o*
	paternal.aunt-1INCL.POSS	paternal.aunt-1INCL.POSS-PL
	'our aunt'	'our aunts'

In the plural forms in (8) and (9), the possessive suffixes *-gu* and *-da* indicate the person and number of the possessor while the plural suffix *-ao* indicates the number of the nominal referent, the possessee.

Body part terms are generally not marked for number. The fact that the body part terms in (10) and (11) have plural referents is clear through context; on the clause level it is also overtly marked through agreement with the subject and object pronoun on the verb respectively. The possessive suffix on the body part noun *nima* 'hand' indicates that the possessor (rather than the possessee) is singular.[5]

(10)	*Nima-na*	*se*	*bado*	*kalili.*
	hand-3SG.POSS	3PL	many	very
	'it has many hands'		Malamala_01BC_246	

(11)	*Nima-na*	*se*	*pai-di.*
	hand-3SG.POSS	3PL	tie-3PL.OBJ
	'they tied his hands'		

In Saliba-Logea, as in other Papuan Tip cluster languages, number is marked on the phrase level if the noun is followed by a modifier. Lexical modifiers and quantifiers agree with their head noun and are obligatorily marked for number by an associative suffix: *-na* if the

[4] For this from only the first syllable is reduplicated: *ta-tau* 'men', possibly to avoid homophony with another lexical item, *tautau* 'picture'.

[5] The subject prefix is written separately from the verb in the Saliba-Logea SIL trial orthography (Oetzel and Oetzel 1997).

referent is singular and *-di* if the referent is plural. The marking of modifiers in this way occurs regardless of whether the noun referent is human.

(12) *numa namwa-namwa-na*
house RED-good-3SG.ASSOC
'good house'

numa namwa-namwa-di
house RED-good-3PL.ASSOC
'good houses'

(13) *lulu posi-posi-na*
shirt RED-white-3SG.ASSOC
'white shirt'

lulu posi-posi-di
shirt RED-white-3PL.ASSOC
'white shirts'

Such number-marked modifiers can also stand alone as the head of a noun phrase:

(14) *laki-laki-na*
big-3SG.ASSOC
'big one'

(15) *loha-loha-di*
RED-long-3PL.ASSOC
'long ones'

The associative suffixes *-na* and *-di* are identical to the singular and plural forms of the third person possessive suffix found on nouns, postpositions, and possessive classifiers, as shown in (16) to (18).[6] This point will be relevant in the discussion in §3 below.

(16) *sina-na*
mother-3SG.POSS
'her mother'

sina-di
mother-3PL.POSS
'their mother'

(17) *sabi-na ye dou*
for-3SG.POSS 3SG cry
'she cried for him'

sabi-di ye dou
for-3PL.POSS 3SG cry
'she cried for them'

(18) *ka-na kai*
CLF-3SG.POSS food
'her food'

ka-di kai
CLF-3PL.POSS food
'their food'

This strategy of referencing number on modifiers and quantifiers is rare in Oceanic languages but is not uncommon in the languages of western Melanesia (Ross 1998, Lynch et al. 2002:40). It is one of the shared morphosyntactic innovations of the Papuan Tip Cluster (cf. Ross 1988:208; Ezard 1997:55; Taubershmidt 1999:14). In some Papuan Tip languages this is the only way to mark number in the noun phrase. The Saliba-Logea constructions belong to type 4 of Ross' (1998:242) overview of possessive-like attributive constructions in the Oceanic languages of north-west Melanesia. Type 4 is the most widespread of the four types. The attributes do not carry nominalising morphology, and in Saliba-Logea they mostly belong to the class of stative verbs. Ross (1998:242–243) observes that many of the languages where Type 4 structures occur are left-branching like Saliba-Logea: they have SOV and Genitive-Noun orders and postpositions. The languages are generally head-marking (in the terminology of Nichols 1986): verbs carry pronominal markers cross-referencing the person and number of the subject and object, and in possessive constructions, it is the possessee which is marked. The possessive-like attributive constructions, which

[6] Since they are identical, distinguishing between possessive and associative suffixes is not straightforward. In this paper I am calling the forms *-na* and *-di* 'possessive suffixes' when they occur on nouns, possessive classifiers and postpositions but 'associative suffixes' when they occur on lexical modifiers and quantifier. There is a full person paradigm of possessive suffixes while associative suffixes only occur in the third person. For further discussion of associative suffixes cf. Lichtenberk (2005), Ross (2001, 1998), and Hooper (1985).

show dependent marking, are the exception to this pattern. Historically, if these constructions are based on possessive constructions, this can be explained through reanalysis of the earlier head (possessed noun) as the dependent (attribute: modifier or quantifier).

Synchronically, Saliba-Logea modifiers carrying the associative suffix differ from possessed nouns in that they are generally reduplicated while the corresponding possessed nouns are not. Compare (19) and (20) with (14) and (15) above.

(19) *laki-na*
big-3SG.POSS
'its size'

(20) *loha-di*
long-3PL.POSS
'their length'

If a human noun with a plural referent is followed by a modifier in Saliba-Logea, then the noun itself carries the plural marker *-ao* while the modifier takes the plural associative suffix *-di*.

(21) *wawaya gagili-na*
child small-3SG.POSS
'small child'

wawaya-o gagili-di
child-PL small-3PL.POSS
'small children'

The situation described so far reflects the standard uses of the Saliba-Logea plural marker and the only contexts where the suffix is obligatory. There are both semantic constraints (only with humans or higher animals) and morphosyntactic constraints on its distribution (only on nouns). Speakers state however that the younger generations tend to use the plural suffix more extensively and in contexts where it may be ungrammatical to the ears of older speakers. As we will see, speakers' perception may be correct with respect to some but clearly not all contexts of use. The remainder of this paper investigates such presumed novel uses, the changes in the selection restrictions and who may be driving them.

3 Extended uses

There are at least two types of extensions to the uses of the plural marker *-ao* described above. One type is semantic: the suffix at times occurs with nouns denoting referents other than humans or higher animals; the other is morphosyntactic: it can occur on words other than nouns. The discussion below draws on a text database of about 16 hours of transcribed spoken language recorded between 1995 and 2006 in the Saliba-Logea area with speakers from a wide age range. Of the data, about 55 percent come from Saliba speakers and about 36 percent from Logea speakers.[7] The transcripts were divided into intonation units and text-audio linked so that the audio is available in text searches and concordances which allows us to verify examples (see Andrew Margetts 2009; Anna Margetts 2009).

3.1 Semantic extension: plural-marking with non-human referents

In terms of the semantic extension, the plural marker occurs on some body part nouns and other relational nouns that are marked as directly possessed, as in (22) to (24) (where the possessor is human).

[7] The remaining nine percent were produced by speakers who were not classified or who were speakers of English.

(22) *Opi-di-**yao*** *se* *duba-duba.*
skin-3PL.POSS-PL 3PL RED-black
'Their skin is black (lit. their skins are black).' Torres_01AC_055

(23) *Mata-di-**yao*** *gibu-gibu-di-yao.*
eye-3PL.POSS-PL RED-blind-3PL.POSS-PL
'They were blind (lit. their eyes are bind).' Boneyawa_03AJ_0012

(24) *Hesa-di-**yao*** *yede* *tem* *wau* *ye* *katai-di* *ne.*
name-3PL.POSS-PL PRSUP DIST.DEM now 3SG pronounce-3PL DET
'Their names are the ones she just mentioned.' Tautolowaiya_02AQ_0137

Morphosyntactically this is a context in which the suffix has always been sanctioned as the affixed word is a noun, but as shown (10) and (11) above, body part nouns are not generally marked for number in this way.

The plural suffix *-ao* also occurs in noun phrases with animal referents which are neither domesticated nor anthropomorphised, as in (25) and (26). (Some of the examples also show aspects of the morphosyntactic extension discussed in §3.2 below.)

(25) *yama* *gagili-di-**yao***
fish small-3PL.POSS-PL
'small fish' Fishing_01BQ_029

(26) *wawaya-**o*** *bigi-bigi-di-**yao***
child-PL RED-big-3PL.POSS-PL
'the big young (of the cuscus)' Gimagima_02AL_045b

Noun phrases denoting plants and parts of plants are also attested with the plural marker, as in (27) to (29).

(27) *taba-taba-di-**yao***
RED-flat-3PL.POSS-PL
'flat ones (leaves)' Daam_01AH_207)

(28) *Lugu-di-**yao*** *loha-loha-di.*
leaf-3PL.POSS-PL RED-long-3PL.POSS
'Their leaves are long.' Bagodu_01AH_075

(29) *kaiwa* *duha-di-**yao*** *udiyedi*
tree hole-3PL.POSS-PL PP.PL
'in tree holes' Gimagima_01AL_018

Finally, in some cases the plural suffix occurs in noun phrases denoting inanimate referents, such as 'style', 'roof' or 'village', as in (30) to (32).

(30) *stail* *hauhau-di-**yao***
style new-3PL.POSS-PL
'new styles' Fishing_01BQ_196

(31) *(yo-di* *numa)* *gatowa-di-**yao***
CLF-3PL.POSS house roof-3PL.POSS-PL
'the roofs (of their houses)' MakingSagoRoof_01AW_036

(32) *magai heka-di-**yao***
village some-3PL.POSS-PL
'some villages' FamilyOrigin_06CL_120

3.2 Morphosyntactic extension: plural marking on other word classes

In its standard use, the plural suffix *-ao* is restricted to nouns, occurring either directly on the stem, or following the possessive suffix if the noun is directly possessed. There are however examples in the data where the suffix occurs on words other than nouns. Examples (33) and (34) show noun phrases where *-ao* is attached to a modifier or quantifier which occurs alone, without a preceding noun.

(33) *Wau heka-di-**yao** se yauyaule.*
PART some-3PL.POSS-PL 3PL crazy
'Some (people) are crazy.' Adoption_01AO_229

(34) *duba-duba-di-**yao***
RED-dark-3PL.POSS-PL
'dark ones (pandanus leaves)' Leiyaha_03AH_120

In other examples the plural suffix occurs on both the head noun and the following modifier, as in (23), repeated here as (35), and (26) above, or it occurs on the modifier but not the head noun, as in (36).

(35) *Mata-di-**yao** gibu-gibu-di-**yao**.*
eye-3PL.POSS-PL RED-blind-3PL.ASSOC-PL
'They are blind (lit. their eyes are bind)' Boneyawa_03AJ_0125

(36) *Siya tamowai bwala-bwala-di-**yao**.*
3PL people RED-lie-3PL.POSS-PL
'They are tricksters (lit. they are lying people).' MalaDoini_01CO_165

The plural suffix is also attested on postpositions, which carry a possessive suffix (which speaks for their nominal origin).

(37) *Ye nuwa-yababa lou-na-o sabi-di-**yao**.*
3SG mind-bad brother-3SG.POSS-PL for-3PL.ASSOC-PL
'She felt bad for her brothers.' Tautolowaiya_01AG_172

(38) *Maiya-na-**o** se miya-miya.*
with.3SG-3SG.POSS-PL 3PL RED-stay
'They lived with her.' TBNatunaoLabui_01CX_0018

Reportedly, some speakers at times also attach the plural suffix to possessive classifiers, as in (39), but such examples are as yet not attested in the database.

(39) *yo-gu-**wao***
CLF-1SG.POSS-PL
'mine (pl)'

In the standard use of the possessive classifier, the number of the possessed object would not be marked (but could be marked on the phrase or clause level through cross-referencing on modifiers or the verb).

3.3 Possessive and associative markers as bridge to extended uses

In the examples discussed above, the morphosyntactically extended uses of the plural marker show the suffix attached to words other than nouns. But, crucially, in all cases these words carry a possessive or associative suffix. It is the presence of such a suffix which can be described as the selection restriction for the plural suffix in its extended uses. In all cases but those involving the postpositions *maiya-* 'with third singular' and *maida-* 'with first plural inclusive' the possessive/associative suffix is the third plural form *-di*.

Interestingly, in examples with plural-marked modifiers or postpositions, the plural marker does not actually add any information. In almost all cases the noun referents are already marked as plural by the associative/possessive suffix *-di*. The only exceptions to this are, again, the postpositions *maiya-* and *maida-* which are attested with possessive suffixes of other person and number distinctions. But they both inherently express plurality of referents: 'participant A together with B' (where A is expressed by the postposition base and B by the referent of the possessive suffix) and so again, the plural suffix does not add any information to the construction.

In examples with modifiers and quantifiers, as in (33) to (37) above, the extended use of the plural suffix results in noun phrases which are double marked for plural by two suffixes, *-di* and *-ao*, both indicating the plural number of the head noun. This is different from (40) to (43), where the plural suffix occurs on part nouns.

(40) *Doha beya-di-**yao**.*
like ear-3PL.POSS-PL
'Like their ears (describing antlers).' FrogStory_01AW_095

(41) *Kinikini-di-**yao** ne hekadi se pulu-pululu.*
young.shoots-3PL.POSS-PL DET some-3PL.POSS 3PL RED-red
'some of the young shoots (of the yams) are red' WekuSinibu_01AC_260

(42) *Doha kaiwa ne lagalaga-di-**yao** ne.*
like tree DET branch-3PL.POSS-PL DET
'Like tree branches.' FrogStory_01AW_098a

(43) *Masi-di-**yao** duba-duba-di ne.*
sap-3PL.POSS-PL RED-dark-3PL.POSS DET
'The ones with the dark sap.' Leiyaha_03AH_128

In these examples, the suffix *-di* marks the plural number of the possessor and the suffix *-ao* marks the plural number of the possessed object and so it does add information to the construction. This is parallel to the marking on kinship terms and possessive classifiers, as in examples (8), (9), and (39) respectively.

Relational nouns (which normally or obligatorily carry a possessive suffix) are some of the most common nouns to be marked by the plural suffix. But in principle, any Saliba-Logea noun can be marked as grammatically possessed. The suffixed nouns in (44) to (46) are not relational and they commonly occur without possessive marking:

(44) *Logea tamowai-na*
Logea person-3SG.POSS
'Logea person'

(45) *yama gulai-na*
fish broth-3SG.POSS
'fish broth'

(46) *Saliba yo Logea waga-di*
Saliba CONJ Logea boats-3PL.POSS
'boats from Saliba and Logea'

In the extended use of the plural suffix, any noun that carries a possessive or associative suffix can be marked by *-ao* and it appears that sometimes the third plural form *-di* is added simply as a vehicle for the plural marker. In a conversation, a Saliba speaker produced the noun in (47) to refer to a strip of shops. The speaker did not specify a referent for the suffix *-di.*

(47) *sitowa-di-**yao***
store-3PL.POSS-PL
'stores' notes07_conversation

The nature of the shops (which could have been specified by an overt referent of *-di* as in *pasa sitowa-di* 'flower shops') was not a topic of the conversation. The purpose of suffixing *-di-yao* seems to have been simply to express the plurality of the shops. Therefore, rather than describing the ongoing grammatical change as a spread of the plural suffix *-ao* it may in fact be more appropriate to talk about the spread of *-di-yao* and analyse this combination as a complex plural suffix which allows the marking of plural number in novel contexts. As mentioned, other person/number distinctions of the possessive suffix are attested only on the postpositions *maiya-* 'with third singular' and *maida-* 'with first plural inclusive'. On modifiers, quantifiers, and possessed nouns only *-di* is attested.

In sum, it seems to be the presence of a possessive or associative marker which provides the morphological bridge that allows the nominal plural suffix *-ao* to spread to other word classes. Relational nouns, postpositions, modifiers, quantifiers and possessive classifiers all provide a possessive or associative suffix to which the plural suffix may, in principle, attach. Associative suffixes on modifiers and quantifiers and their number agreement with the head noun is a feature particular to Papuan Tip cluster languages, and some other Western Melanesian languages and may, in part, be responsible for the spread of the plural suffix in Saliba-Logea.

3.4 Distribution of examples

The text database shows only few examples of the plural marker's extended use: 79 instances in over 24,600 intonation units and about 16 hours of recordings. The low numbers may partly be due to the fact that the novel uses are likely to be more common in the casual context of conversation and that our database includes more narratives and procedural texts.

In this section I provide an overview and breakdown of the types of examples, and the age and dialect of the speakers.

3.4.1 Age group and dialect

As mentioned above, older Saliba speakers sometimes bemoan the younger generations' use of the plural marker in contexts where it sounds inappropriate to them. It was statements of this kind that raised my interest in the marker's distribution and usage extension.

Investigating differences in the linguistic behaviour across different cohorts of speakers requires some methodological decisions. Speakers could be grouped, for example, based on their age at the time of recording, or based on their date of birth. Given that the database used for this study includes recordings from a time span of more than ten years (1995 to 2006), the first method would result in the same speaker showing up in different age groups if they contributed to the database in more than one year. Therefore, in order to investigate the use of the plural marker across different cohorts, speakers were assigned to 'generations' based on their approximate birth year. The generations were defined in steps of 15 years. Each speaker was assigned to a generation and the generation label (A, B, C, D, or E) was included in the Toolbox concordance so that examples of the plural marker could be sorted by generation.[8]

Based on speakers' comments on young people's use of the plural marker, an age split in the speakers who produced extended-use examples could be expected, with the younger generation producing more of the examples. However, this turned out not to be the case. There are a total of 27 speakers who produced the 79 extended-use examples. Over 60 percent of examples were produced by speakers belonging to the two oldest generations (who were 54 or older in 2009) and only 12 percent where produced by speakers of the two youngest generations (who were 38 or younger). The breakdown of examples by generation is provided in Table 1.

Table 1: Breakdown of extended plural examples by generation

	Total	%
Total	**79**	**100%**
Gen A (*before 1940)	24	30%
Gen B (*1941–1955)	30	38%
Gen C (*1956–1970)	16	20%
Gen D (*1971–1985)	6	8%
Gen E (*1986–2000)	3	4%

Sorting the data by speaker revealed however that the majority of examples came from speakers of the Saliba dialect. Only 14 of the 79 examples (18%) were produced by Logea speakers even though they produced about 36% of the recordings in the corpus (measured in intonation units).

Table 2 gives the breakdown of extended-use examples of the plural suffix by generation and dialect.

[8] This methodology does not allow us to investigate the possibility of a linguistic feature being used more extensively across all or some generations at different points in time. Therefore this study did not investigate whether, overall, speakers produced more extended-use examples in, say, 2005 than in 1995 (but impressionistically this is not the case).

Table 2: Breakdown of extended plural examples by generation and dialect

	Total	Saliba		Logea	
Total	**79**	**65**	**82%**	**14**	**18%**
Gen A (*before 1940)	**24**	23	96%	1	4%
Gen B (*1941–1955)	**30**	26	87%	4	13%
Gen C (*1956–1970)	**16**	7	44%	9	56%
Gen D (*1971–1985)	**6**	6	100%	–	–
Gen E (*1986–2000)	**3**	3	100%	–	–

These figures are not easy to interpret as the distribution of speakers of the two dialects is not balanced across the different generations in the database. In fact, the majority of intonation units in Saliba were produced by speakers of generations A and B while the majority of intonation units in Logea were produced by speakers of generations C and D.[9]

To better interpret the figures in Table 2 it would be helpful to compare these examples of extended uses of the plural markers to the number of cases where it *could* have been used, and where it would have counted as an example of extended use, but where it did *not* occur. At this stage it is impossible to produce figures for the number of plural noun phrases with non-human referents in the database which are not marked by the plural suffix. There is no search function or filter which could produce such figures.[10] However, the examples of extended uses of the plural suffix all follow a possessive or associative suffix. Therefore, instances of these suffixes on words other than human nouns and without a following plural suffix may serve as a reasonable base of comparison here. There are a total of 763 of such tokens. This figure excludes human nouns, the possessive classifiers *yo-* and *ka-* and some other high-frequency items which occur in fixed expressions and which are not attested with the plural suffix. Excluded were also a number of possessed nouns which are overtly marked for singular number by the singular indefinite marker *hesau* 'a/other' and could therefore not take the plural suffix *-ao*.[11] The 763 instances form the group of examples which, if they *were* marked by the plural suffix would constitute examples of extended use of the plural suffix. In the tables below, I will refer to this set as the 'non-plural' examples.

Based on this, the figures in Table 2 can be evaluated more meaningfully. Combining the non-plural examples with the examples of extended use provides us with the number of potential extended-use examples (i.e. those that could have been marked by the plural suffix but are not, plus those that are marked). On this basis it can be calculated what percentage of words which could in theory be marked by the plural suffix in its extended use were actually marked. Table 3 presents the total figures and a breakdown of examples

[9] There are a number of reasons for the differences in the representation of the dialects in the database. One is that recording of Saliba data started in 1995 while Logea data was only recorded from 2005 when fewer speakers of the older generation were still alive.

[10] The extended-use examples of the plural suffix were collected from the entire database but only part of the database is interlinearised. This restricts the parameters which can be used in searches.

[11] As a reminder, words marked with the third person plural possessive/associative suffix *-di* do not necessarily have plural referents. On nouns, the suffix marks the plural number of the possessor rather than of the referent itself.

by generation. The second last column shows the total number of items which have the potential to be marked, and their distribution in tokens (and in percentages) across the generations. The last column provides the percentage of plural-marked examples.

Table 3: Potential and actual extended uses of PL suffix by generation

	non-pl	with pl	Total	% with pl
Total	**763**	**79**	**842 (100%)**	**9%**
Gen A (*before 1940)	253	24	277 (33%)	**9%**
Gen B (*1941–1955)	102	30	132 (16%)	**23%**
Gen C (*1956–1970)	254	16	270 (32%)	**6%**
Gen D (*1971–1985)	146	6	152 (18%)	**4%**
Gen E (*1986–2000)	8	3	11 (1%)	**27%**

Generation E (age 9 to 23 in 2009) shows the highest percentage of extended uses of the plural suffix but it has to be noted that for this generation there are only very few examples overall (they constitute only 1% of potential examples). Speakers marked three out of 11 possibilities (27%). Speakers of generation B (age 54 to 68 in 2009) choose to mark an 'eligible' word in 23 percent of possible contexts. Speakers of generation A choose to add the plural suffix in nine percent of possible case. Generations C and D choose to mark six and four percent respectively. The table shows that the extended use of the plural suffix in novel contexts is not clearly driven by the younger generations as would have been expected based on speakers' comments. (There may still be aspects of the extension which are driven by the younger generations. An example may be the use of the plural marker on possessive classifiers, which speakers have quoted as an inappropriate use, which is however not attested in the text database so far.)

To evaluate the number of *speakers* who produced extended-use examples, we can consider the total number of speakers of each dialect who produced words that *could* be marked by the plural suffix and calculate the percentage of speakers who did in fact use the plural suffix in one of the novel contexts. There are 31 Saliba speakers who produced words that could be marked and 19 of them (61%) used the plural suffix at least once. Of Logea speakers there are a total of 19 who produced words that could be marked and eight of them (42%) used the plural suffix at least once.

Table 4, parallel to Table 3 above, provides the breakdown of tokens for the potential and actual extended uses in the two dialects. The last column shows the percentage of plural-marked examples.

Table 4: Potential and actual extended uses of PL suffix by dialect

	non-pl	with pl	Total	% with pl
Total	**763**	**79**	**842**	**9%**
Saliba	339	65	404	**16%**
Logea	424	14	438	**3%**

The table shows that Logea speakers choose to use the plural suffix in the discussed novel contexts in only three percent of possible cases; and much less frequently than Saliba speakers who choose the plural suffix in 16 percent of possible contexts.

It appears, therefore, that the extended use of the plural marker is driven by Saliba speakers and more established in the Saliba dialect. (In fact, a Logea speaker stated that the addition of *-yao* or *-di-yao* on modifiers and quantifiers sounds like the Saliba dialect to her ears.) This finding is interesting in that there are otherwise no established grammatical differences between the dialects.

3.4.2 Syntactic and semantic types

Since there is a semantic and a morphosyntactic aspect to the use extension of the plural marker it is relevant to investigate the frequency of the different types of attested examples. Table 5 shows the breakdown of the 79 examples by the semantics of the referent and by constituent type.

Table 5: Semantics of noun referents across different plural-marked word classes

	Total	Noun referent			
	79	human (+body parts)	animal	plant (+plant parts)	other
modifier	14	3	5	2	4
quantifier	29	12	3	1	13
postposition	12	12	–	–	–
		Possessor referent			
		human	animal	plant	other
possessed noun	24	3	3	9	9

In the largest group of examples (29 of 79) the plural suffix appears on a quantifier. In all but one case the quantifier is *heka-di* 'some'. In some cases the quantified noun directly precedes the quantifier, as in (48) and (49). In other instances *heka-di* stands alone and the noun occurs somewhere in the preceding discourse, as in (50) and (51).[12]

(48) *tautau-na* *maiya-na* *sikulu* *heka-di-**yao***
picture-3SG.POSS with.3pers-3SG.POSS school some-3PL.POSS-PL
'a picture of her with some school children' MangoTree_01AH_0070

(49) *ye-di* *laugagayo* *heka-di-**yo***
CLF-3PL.POSS law/advice some-3PL.POSS-PL
'some of their laws' Customs_01DS_0032

(50) *heka-di-**yao*** *se* *yauyaule*
some-3PL.POSS-PL 3PL crazy
'some of them are crazy' Adoption_01AO_0299

[12] The quantifier *heka-di* 'some' is attested with two different allomorphs of the plural suffix. There are ten instances with *-yao* and eighteen instances with *-yo*. There are no clear factors determining this choice but individual speakers seem to use either one or the other, but not both.

(51) *heka-di-**yo*** *nuwa-gu* *se* *luluhi-di*
some-3PL.POSS-PL mind-1SG.POSS 3PL forget-3PL.OBJ
'I forgot some of them' TBlaki_02AT_0012

By way of comparison, there are 87 instances of *heka-di* 'some' without the plural suffix in the database.

In one instance the plural suffix appears on the quantifier *bado-di* which is used when counting entities as in (52). (As a stative verb *bado* expresses the concept 'to be many'.)

(52) *buluka* *bado-di-**yao*** *five o four*
pig many/number-3PL.POSS-PL five or four
'five or four pigs' Beyabeyana_02CZ_0094

For comparison, there are 16 instances of *bado-di* without the plural suffix.

In the second largest group of examples (24 of 79) the plural suffix appears on possessed nouns. This category includes part nouns (body parts, plant and house parts) and referents which are more loosely associated with one another than part-whole relations (e.g. baskets of food, scoops of salt). Some examples are provided in (53) to (55).[13]

(53) *mata-di-**yao*** *wa* *se* *he-laki-di*
eye-3PL.POSS-PL ANA 3PL CAUS-big -3PL.POSS
'they opened their eyes wide (in astonishment)' Mouse7_02DA_0034

(54) *doha kaiwa ne* *lagalaga-di-**yao*** *ne*
like tree DET branch-3PL.POSS-PL DET
'like tree branches' FrogStory_01AW_0099

(55) *giyahi* *numa-di-**yao*** *udiyedi*
feast house-3PL.POSS-PL PP.PL
'in the feast houses' Beyabeyana_02CZ_0029

Of the total 79 examples 14 show the plural suffix on lexical modifiers, such as in (56) and (57).

(56) *giyahi* *laki-laki-di-**yao***
feast RED-big-3PL.POSS-PL
'big feasts' Beyabeyana_02CZ_0200

(57) *bigisipi-di-**yao*** *ne*
big-3PL.POSS-PL DET
'big ones (cuscus)' Gimagima_02AL_0073

The remaining twelve examples show the plural suffix on postpositions. One example, presented in (37) above, shows the plural suffix on the postposition *sabi* 'for'. In nine cases the postposition is *maiya-* 'with third singular', as in (58) and (59), and in two cases it is *maida-* 'with first plural inclusive', as in (60).

(58) *maiya-da-**o*** *ta* *lao* *ta* *paisowa*
with.3SG-1INCL.POSS-PL 1INCL go 1INCL work
'(tell them) to go and work with us' Tautolowaiya_01AG_0040

---

[13] I cannot provide figures for possessed nouns or lexical modifiers which do *not* carry the plural marker for comparison. This would require a concordance sorting on an item of the part-of-speech line but this only exists for the subpart of the database which is intelinearised at this point and could therefore not produce meaningful figures.

(59) *maiya-di-**yao*** *se* *kai-kai*
with.3SG-3PL.POSS-PL 3PL RED-eat
'they eat with them' HairCutting_04AA_0016

(60) *Moli* *kabo* *maida-gu-**wao*** *ka* *lao-ma*
Molly will with.1EX-1SG.POSS-PL 1EX go-hither
'I'll come with Molly' Fishing_01BQ_0578

For comparison, there are 67 instances of *maiya-* 'with third singular' without the plural suffix and 41 instances of *maida-* 'with first plural inclusive'. Both postpositions always express plurality of participants, whether or not the plural suffix is present.

Tables 6 and 7 show the types of examples attested across age groups for the two dialects.

Table 6: Breakdown of Saliba examples by generation and marked constituent

	Total	Modifier	Quantifier	Postposition	Possessed noun
Total	**63**	**11**	**20**	**8**	**24**
Gen A (*before 1940)	**23**	1	6	7	9
Gen B (*1941–1955)	**24**	8	10	–	6
Gen C (*1956–1970)	**7**	1	4	–	2
Gen D (*1971–1985)	**6**	–	–	1	5
Gen E (*1986–2000)	**3**	1	–	–	2

Table 7: Breakdown of Logea examples by generation and marked constituent

	Total	Modifier	Quantifier	Postposition	Possessed noun
Total	**14**	**1**	**9**	**4**	–
Gen A (*before 1940)	1	–	1	–	–
Gen B (*1941–1955)	4	1	2	1	–
Gen C (*1956–1970)	9	–	6	3	–
Gen D (*1971–1985)	–	–	–	–	–
Gen E (*1986–2000)	–	–	–	–	–

The tables show that while examples of possessed nouns occur across all five generations in Saliba they are totally absent in the Logea data set. In both dialects the quantifier *heka-di* 'some' is among the most frequently marked items.

The 14 Logea examples were produced by eight speakers and they involve only three different lexical items: *hekadi* 'some', *maiyana* 'he/she with her/him', and a modifier, *bwalabwaladi* 'lying/tricking (ones)'. The 65 Saliba examples were produced by 19 speakers and they involve 32 different lexical items.

4 Conclusion

As in other Oceanic languages, plural marking in Saliba-Logea seems to have been originally restricted to a small class of essentially human nouns. But the selection criteria widened and the plural suffix *-ao* is nowadays attested in noun phrases with non-human

referents, on part nouns with human and non-human possessors and on members of other word classes. One can hypothesise a bridge by which the plural marker extended from human relational nouns, such as kinship terms, which carry a possessive suffix, to relational nouns with non-human referents and then to non-relational nouns and other word classes which can carry the same possessive/associative marker. Such markers occur not only on nouns but also on modifiers, possessive classifiers, postpositions and possessive classifiers. (However, the presence of the possessive/associative morphology suggests a nominal origin for all of these classes.) The occurrence of such markers on lexical modifiers and some quantifiers to mark number agreement with the head noun is one of the shared innovations of the Papuan Tip Cluster (Ross 1988:208) and seems to be restricted to Western Melanesian languages (Lynch et al. 2002:40). It is in fact the combination of the third person plural possessive/associative suffix *-di* and the plural suffix *-ao* which allows plural marking in most of the novel contexts and the sequence *-di-yao* can be analysed as a morphologically complex plural marker. It seems that, nowadays, essentially any noun can take the plural marker *-di-yao* independent of its referent.

There is no clear split between older and younger speakers in the use of the plural marker. It is the youngest and the second oldest speaker cohort that produced the highest percentage of examples in this study. However, the extended use of the plural maker is clearly more common for speakers of the Saliba dialect than for Logea speakers. This is of special interest as there are otherwise virtually no grammatical differences between the dialects. Although this has not been investigated in this paper, it is possible that the spread of the plural marker is partly motivated by analogy with English, the lingua franca of Milne Bay Province and by the speakers' attempts to mark plurality more explicitly than Saliba-Logea grammar originally allowed.

References

Abel, Cecil. n.d. Suau (grammatical notes). Unpublished manuscript.

Ezard, Bryan. 1997. *A grammar of Tawala: an Austronesian language of the Milne Bay area, Papua New Guinea*. Canberra: Pacific Linguistics.

Hooper, Robin. 1985. Proto-Oceanic *qi. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 141–167. Canberra: Pacific Linguistics.

Lichtenberk, Frantisek. 1982. Individuation hierarchies in Manam. In Paul Hopper and Sandra Thompson, eds *Studies in transitivity*. Syntax and Semantics 15, 261–276. New York/San Francisco/London: Academic Press.

—— 2005. On the notion of 'adjective' in Toqabaqita. *Oceanic Linguistics* 44(1).113–144.

Lithgow, David. 1975. A grammatical analysis of a Dobu text. In Richard Loving, ed. *Papers in five Austronesian languages, workpapers in Papua New Guinea languages, 12*, 25–56. Ukarumpa: Summer Institute of Linguistics.

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

Margetts, Andrew. 2009. Using Toolbox with media files: a guide to the processes involved. *Language Documentation and Conservation* 3(1).51–86.

Margetts, Anna. 1999. *Valence and transitivity in Saliba, an Oceanic language of Papua New Guinea*. MPI Series in Psycholinguistics. Nijmegen: Max Planck Institute for Psycholinguistics.

—— 2009. Data processing and its impact on linguistic analysis. *Language Documentation and conservation*. 3(1).87–99.

Mosel, Ulrike. 1994. *Saliba*. Languages of the world, Materials, 31. München/Newcastle: LINCOM Europa.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62(1).56–119.

Oetzel, Rainer, and Sabine Oetzel. 1997. Orthography and phonology description of Saliba. Unpublished manuscript. Ukarumpa: SIL.

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

—— 1998. Possessive-Like Attribute Constructions in the Oceanic Languages of Northwest Melanesia. *Oceanic Linguistics* 37(2).234–276.

—— 2001. Proto-Oceanic *i, *qi, and *-ki. In Joel Bradshaw and Kenneth L. Rehg, eds *Issues in Austronesian morphology. A focusschrift for Byron W. Bender*, 259–278. Canberra: Pacific Linguistics.

Taubershmidt, Gerhard. 1999. *A grammar of Sinaugoro, an Austronesian language of the Central Province of Papua New Guinea*. Canberra: Pacific Linguistics.

22 *Verbs of perception in Proto Oceanic*

MEREDITH OSMOND AND ANDREW PAWLEY

1 Introduction[1]

This paper investigates verbs of perception in Proto Oceanic, based on a comparison of a sample of daughter languages. A full comparative study of the morphology, syntax and semantics of this set of verbs in Oceanic languages would require a book. Here we offer an introductory account, focusing mainly on certain basic semantic and grammatical features of perception verbs, and building on the work of Bethwyn Evans (2003), whose study of verb classes and valency-changing devices in Proto Oceanic includes a section on several verbs of perception.

Since Aristotle, Western scholars have generally assumed that humans have five basic senses: sight, hearing, smell, taste and feeling by touch. In a basic sensing event there is an animate participant, the experiencer, who by means of a body part (eye, ear, nose, tongue, skin) becomes aware of a separate participant (the stimulus or source). Neurophysiological research shows that the five senses scheme is too simple. People have additional physiological systems for sensing pain, temperature, balance and awareness of how our body and limbs are moving (proprioception).

Languages of the world generally give these non-basic senses different grammatical treatment from the basic senses. There are a number of possible explanations for this. Firstly, no readily-recognised sense organs are participants in sensations that come through these other physiological systems. Secondly, sensations such as pain, dizziness, and feeling cold are involuntary, whereas in the case of seeing, hearing, smelling etc. the experiencer may initiate the process and at least has a measure of control over it. Third, whereas the stimulus or source of a basic sensory experience is typically an identifiable entity outside the experiencer's own body (the thing seen, heard etc.) the source of non-basic sensations like pain, cold or dizziness is not external and may not be identifiable. Because the

[1] We offer this paper as a small token of our esteem for Malcolm, friend and long-time colleague in Oceanic and Papuan comparative projects. We have additional reason to be grateful to Malcolm. Under the impression that this paper was intended for inclusion in a forthcoming Lexicon of Proto Oceanic volume, Malcolm has suggested a substantial number of amendments and additions to an earlier truncated version of this paper which have now been incorporated. We hope this version does not cause him too much angst. We are also grateful to two anonymous reviewers who made a number of perceptive comments.

sensations may be prolonged, they are often treated as states and the focus tends to be on their effect on the body. Thus in English we typically describe feeling pain, fear, cold, itchiness and dizziness in terms of the experiencer or a body-part being in a state or condition, which is expressed by a predicate adjective ('My hip is quite painful', 'Are you cold?', 'Mary is dizzy') whereas for the primary sensing events we tend to use active verbs/verbs with the experiencer as actor ('I saw/heard John'), rather than adjectival predicates with the experiencer as involuntary recipient of the stimulus ('John is visible/ audible to me').

The present paper will deal mainly with the treatment of the five basic senses in Oceanic languages and with the question of whether the different senses receive similar grammatical and semantic treatment.

Basic perception verbs vary conceptually along a number of parameters. These are illustrated in English in the following paradigm, closely based on that proposed by Viberg (1984). We have labelled the variables as (i) sensing, (ii) attending and (iii) stimulus-subject.[2]

Table 1: English perception verbs
(based on Viberg's basic paradigm of verbs of perception)

Sense modality	Sensing	Attending	Stimulus-subject
sight	I see many people	I look at the film	The film is visible/looks blurry.
hearing	I hear bells	I listen to the tune	The tune is audible/sounds loud.
smell	I smell smoke	I smell the milk	The milk is smelt?/smells sour.
taste	I taste garlic	I taste the mixture	The mixture is tasted/tastes fine.
touch	I feel the wind in my hair	I feel the fabric	The fabric is felt?/feels velvety.

Some languages distinguish lexically or grammatically between two kinds of perception events involving the basic senses: *sensing* and *attending*. A verb or verbal clause depicting a *sensing* event focuses on the animate participant's experiencing of the stimulus; it is neutral as to whether this experience was intentional or accidental. In the case of an *attending* event, by contrast, an experiencer is depicted as intentionally focusing on a target. Languages may express the difference lexically, as is done in the English verbs *see* versus *look* and *hear* versus *listen*, but not in the verbs *smell*, *taste* or *feel*, where the same term can be used for both sensing and attending. In either event the experiencer will be subject of the verb, and the verb will usually be transitive. A matter to be investigated is the way in which intent is signalled in the basic sensory verbs in Oceanic languages, and its corollary, whether, for each of the basic senses, a language will use the same verb for both sensing and attending events.

We may define a canonical perception verb (and clause) cross-linguistically as one that has the perceiver (experiencer) as the highest ranked argument (the subject in nominative-accusative languages). However, it is common to find other kinds of clauses used to represent perceptions. When the focus shifts from the performance of the act to some

[2] Viberg uses the following labels: Experience (= sensing), Activity (= attending) and Copulative (= stimulus-subject).

conclusion, the source of the perception, the stimulus, will be subject and the verb will be intransitive. Focus may then, at least for sight and hearing, be limited to acknowledgement of the perception 'it is seen/it is heard' or even acknowledgement of the ability to be perceived 'it is visible/audible'. More commonly, further information may be given by a qualifier in the case of all five senses 'it looks fine/it sounds awful/it smells sour etc.'. The degree to which languages use the same verb polysemously varies widely. In English, for example, *smell*, *taste* and *feel* can all be used with experiencer or stimulus as subject, while *see* and *hear* may use a related or different lexeme for stimulus-subject.

When dealing with an intransitive verb in many Oceanic languages, one must ask: Is this verb active or stative? Does it take as subject (or highest-ranked argument) an Actor or an Undergoer? The intransitive forms of many verbs of process or change of state, such as those that mean 'open', 'close', 'break', 'cut', 'split', 'burn' and 'block' are typically stative, taking as subject the thing that undergoes the process. Many intransitive verbs, both active and stative, can be transitivised by adding (a) a transitive suffix of the form *-i* or *-(C)i* (where C is a variable consonant) and (b) an object pronoun suffix or clitic; or simply by adding (b).

Sometimes a language will use a single perception verb form polysemously to represent two or three basic senses and sometimes also to represent cognitive processes like knowing, thinking, understanding and remembering, and cultural practices like obeying, paying attention and learning. Given that sensory verbs are often polysemous in these ways, the question arises whether there is a universal hierarchy within which senses are ordered, which will predict the direction of semantic extension. Viberg (1984) finds some evidence for the following hierarchy: sight > hearing > touch > smell, taste. This hierarchy implies that vision has primacy over the other senses, such that a verb of seeing may be extended to refer to at least certain senses lower on the scale, but not the reverse. Hearing in turn has primacy over touch, smell and taste.

With these issues in mind, let us turn to the Oceanic languages.[3] Our project is hampered by the fact that dictionaries and grammars of Oceanic languages seldom provide careful and detailed descriptions of the grammar and semantics of verbs of perception. Dictionary entries often fail to state whether a particular verb is transitive or intransitive, and derived forms are often not given full glosses. In some cases these gaps in the data limit our ability to make secure reconstructions.

2 Oceanic cognate sets

2.1 Verbs of seeing

All Oceanic languages have at least one transitive verb whose primary sense is 'see s.t.' (and which may also mean 'look at s.t.'). The experiencer is the subject and the source/stimulus the direct object. Typically they also have a number of transitive verbs for intentional visual activities comparable, for example, to English 'peer', 'peep', 'glance', 'gaze' and 'stare'. Verbs which we might call 'verbs of directed looking' (*look around, up, down, in, out, over, away, into* etc.) are likely to be expressed in Oceanic languages by a serial verb construction or by a combination of verb and directional marker (cf. Ross 2003:256, 2004).

---

[3] Data sources are as listed in volume 2 of *The lexicon of Proto Oceanic*, Appendix 1. *ACD* is the Austronesian Comparative Dictionary, compiled by Robert Blust. *POLLEX* is Clark and Biggs' Polynesian Lexicon. Both are computer files.

Verbs of seeing and looking may also be used intransitively with the perceiver as subject. Examples are Motu *ita* (VI) 'see, look', *ita-ia* (VT) 'look at s.t.', Arosi *rio* (VI) 'see', *rio-si* (VT) 'look at s.t.', and Wayan Fijian *tola* 'see, look', *tola-vi-* 'see s.t.'.

POc *kita (VI) 'see', *kita- (VT) 'see s.t.' is a well-supported reconstruction with reflexes in both Western Oceanic and Eastern Oceanic languages. It is also noteworthy that a number of Eastern Oceanic languages have extended the meaning of their reflexes to include 'know' and 'understand'.[4]

PMP *kita 'see' (Dempwolff)[5]

POc *kita (VI) 'see', *kita-i- (VT) 'see s.t.'

NNG:	Tuam	*(i)gita*	'see'
NNG:	Malai	*(i)gita*	'see'
NNG:	Matukar	*ita*	'see'
NNG:	Manam	*ita*	'see, look at'
PT:	Gumawana	*gita*	(VI) 'see'
		gite-	(VT) 'see s.t.' (*-a > e, assimilation)
PT:	Dobu	*ʔita*	(VT) 'see, look'
PT:	Motu	*ita*	(VI) 'see, look'
		itai-	(VT) 'look at s.t.'
MM:	Meramera	*ite*	'see' (*-a > e, assimilation)
NCV:	Lo-Toga	*itɛ*	'see'
NCV:	Raga	*ɣita*	'see'
SV:	Kwamera	*ata* (alt. *ati*)	(VI, VT) 'see, look, regard, understand'
SV:	Anejom	*(e)ɣet*	'see'
Mic:	Carolinian	*gɨtt*	'look for, search'

PPn *kite 'see, appear, know' (*-a > e, assimilation)

Pn:	Tongan	*kite*	(VSt) '(of distant objects) to appear, be or come in sight'
Pn:	Niuean	*kite*	(VT) 'see, learn, understand, know'
Pn:	E Futunan	*kite*	(VSt) 'appear in distance, be seen'
Pn:	Rennellese	*kite*	'look, see, find'
Pn:	Tikopia	*kite*	'see, look at, catch sight of'
Pn:	Tahitian	*ʔite*	'see, know, recognise'
Pn:	Marquesan	*kite*	'recognise, see, know'
Pn:	Maori	*kite-*	'see, find'

---

[4] Some Proto Central Pacific languages identify a 'sixth sense'. PCP *ki[t,d]a-vi (VT) 'to sense without actually seeing, hearing, etc., have a premonition that s.t. will happen', is reconstructable based on reflexes in Bauan and Wayan Fijian, Tongan, Samoan, Maori and Rarotongan. This is doubtless cognate with POc *kita 'to see'.

[5] Abbreviations for protolanguages and language groups are: Adm - Admiralties; Fij - Fijian; Mic - Micronesian; MM - Meso-Melanesian; NCal - New Caledonian; NCV - North/Central Vanuatu; NNG - North New Guinea; PAn - Proto Austronesian; PCP - Proto Central Pacific; PEOc - Proto Eastern Oceanic; PMic - Proto Micronesian; PMP - Proto Malayo-Polynesian; Pn - Polynesian; PNCV - Proto North/Central Vanuatu; POc - Proto Oceanic; PPn - Proto Polynesian; PSS - Proto Southeast Solomonic; PT - Papuan Tip; PWMP - Proto Western Malayo-Polynesian; PWOc - Proto Western Oceanic; SES - Southeast Solomonic; SHWNG - South Halmahera-West New Guinea; SV - South Vanuatu; WMP - Western Malayo-Polynesian.

There are a number of competing reconstructions with some claim to be the general term for 'see'. Reflexes of *kita and *reki[-] (with doublet *reqi[-]) occur almost in complementary distribution (with some overlap in North New Guinea), and a distinction in meaning between them cannot be clearly identified. Both *reki[-] and *reqi[-] are reconstructable to POc with no clear difference in meaning. Only Bugotu and Gela reflect both members of this pair with the reflex of *reki[-] referring to seeing and the reflex of *reqi to directed looking.

POc *reki[-], *reqi[-] 'see, look, see s.t., look at s.t.'

NNG:	Maleu	*lei*	'see'
NNG:	Mangap-Mb.	*re*	(VT) 'see, look, experience; consider, think, be aware'
NNG:	Yabem	*liʔ*	'see, look at s.t., know, have experience'
NNG:	Hote	*ye*	'see'
NNG:	Lamogai-R	*rike*	'see'
NNG:	Amara	*rei*	'see'
MM:	Bilur	*re*	'see'
MM:	Siar	*re*	'see'
MM:	Banoni	*reɣe*	'see'
MM:	Babatana	*ri*	'see'
SES:	Bugotu	*reɣi*	(VT) 'see'
		rei	'to look'
SES:	Gela	*riɣi*	(VT) 'see' (*riɣi sondo* 'to find', *riɣi puku* 'see clearly', *riɣitaoni* 'look after, take care of') 'see, look' (in compounds meaning 'look up/about/here/round, stare at, squint' etc)
SES:	Lau	*riki-a*	'see'
SES:	Toqabaqita	*riki-a*	(VT) 'see, look at, watch'
SES:	Arosi	*rē-i*	'see'
SES:	Fagani	*riɣi-a*	'see'
SES:	Bauro	*reɣi-a*	'see'
SES:	Kahlua	*reɣi-a*	'see'
Fij:	Rotuman	*rāe*	'see, espy, catch sight of, find'
Fij:	Bauan	*rai-ða*	(VT) 'see s.t.'

Reflexes of *liqos also suggest that its POc meaning referred to directed looking.

POc *liqos (VI) 'look, see', *liqos-i- (VT) 'look at s.t., see s.t.'

MM:	Nakanai	*liho*	'to see, look at'
SES:	Bugotu	*liohi-*	(VT) 'look at s.t.'
SES:	'Are'are	*rio*	'see, look, be awake' (in many compounds: 'look for, around' etc)
SES:	Toqabaqita	*lio*	(VI) 'look, look after'
		lio(nūna)	(VT) 'look at oneself (as in a mirror)'
SES:	Kwaio	*lia*	(VI) 'see, look' (*o > a irregular)
		liasi-	(VT) 'see s.t.'
SES:	Sa'a	*lio, lio-lio*	(VI) 'to look, see, be awake'
SES:	Ulawa	*liosi-*	(VT) 'see s.t.'
SES:	Arosi	*rio*	(VI) 'look, see'
		riosi-	(VT) 'look at s.t.'

PNCV *leʔo-si ‘see, look at’ (Clark 2009)

NCV:	Araki	*lesi*	‘see’
NCV:	Uripiv	*(e-)lesi-*	(VT) ‘see, look at s.t.’
NCV:	Paamese	*lesi-*	(VT) ‘see, look at s.t.’

PSV *(e-)laqVs ‘look at, look for’ (Lynch 2001)

SV:	Anejom	*(e-)laθ*	‘look in certain direction’
SV:	Sye	*ela(saɣ)*	‘look up’
		ela(mpya)	‘look away’

PMic *lō, *lō-Si ‘see’ (Bender et al.)

Mic:	Kiribati	*nō*	‘look on’
Mic:	Marshallese	*lew*	‘see’
		lew-ey	‘see s.t.’

Another putative POc reconstruction, *ta(d,dr)aq has reflexes in a number of Meso-Melanesian languages that predominantly mean ‘see’. In other subgroups its reflexes more often mean ‘look at’, ‘observe’ (Micronesian) or to ‘look upwards’ (North New Guinea and South East Solomonic). If POc *ta(d,dr)aq proves to be related to PMP *tiŋadaq ‘look up, look skyward’, reconstructed by Dempwolff (1938), it would support the ‘look upwards’ gloss.

POc *ta(d,dr)aq (VI) ‘look, look up’, *ta(d,dr)aq-i (VT) ‘see s.t., look up at s.t.’

Adm:	Mussau	*tara*	‘to look’
		tara(kila)	‘recognise’ (*kila* ‘know (people)’)
Adm:	Tenis	*tara(ie)*	‘see’
NNG:	Manam	*tada*	(VI) ‘look up’
		tada(li)	(VT) ‘look up to s.o., s.t.’
MM:	Tigak	*tara(i)*	‘see’
MM:	Solos	*tara*	‘see’
MM:	Halia	*tara*	‘see, look’
MM:	Selau	*tara*	‘see’
MM:	Teop	*tara*	‘see’

PSS *tada, tadaq-i- ‘look at s.t., look up to s.t.’

SES:	Bugotu	*tada*	‘look up’
SES:	Gela	*tada*	‘face up, upwards’ (*tada-tada* (VT) ‘look up’)
SES:	Lau	*ada*	(VI) ‘to open the eyes, use the eyes; see, look’
SES:	Kwai	*ada*	‘see’
SES:	Arosi	*āda*	(VI) ‘look up, raise the eyes’
		ādaʔi	(VT) ‘look up to’
SES:	Bauro	*ata*	‘look up’
Mic:	Carolinian	*saeṣēy*	(VT) ‘look for s.t or s.o., look at or observe s.t.’ (respect)
Mic:	Woleaian	*saṣēy*	(VT) ‘look at s.t., observe s.t.’
cf. also:			
MM:	Vitu	*ɣada*	‘see’
MM:	Lavongai	*ara(i)*	‘see’

Proto Oceanic also had a number of lexemes of visual perception carrying additional information as to manner, duration, purpose etc. We have reconstructed POc *tirop, *tirop-i ‘look intently’, *kilop, *kilop-i ‘glance, glimpse’, *kila(t) (VI) ‘see clearly, discern, recognise’, and *(s,j)ila(k) ‘look sideways, glance around’. Similarities of form between *tirop, *kilop and *kila(t) may have led to some crossover of meaning in reflexes.

PMP *tin[d]ap ‘look intently’ (Dempwolff)

POc *tirop (VI), ‘look intently, as at reflection or searching for lice’; *tirop-i- (VT) ‘look at s.t., look for s.t. intently’

NNG:	Gitua	*tiro*	‘look for’
NNG:	Medebur	*(i)tir(to)*	‘look for’
NNG:	Wogeo	*(i-ti)tiri*	‘look for’
MM:	Roviana	*ti-tiro*	‘search for’
		tiro	‘to read’
		tiro(ana)	‘a mirror’
SES:	Bugotu	*tiro*	(V) ‘to look’; (N) ‘a pool, window glass, mirror’
SES:	Toqabaqita	*iro*	(VI) ‘look for s.t., search’ (takes an oblique object)
		iro-a	(VT) ‘look or search for s.o., s.t.’
SES:	Lau	*iro*	(VI) ‘look’
		irofi-	(VT) ‘look at s.t. fixedly, look for s.t.’
SES:	Kwaio	*ilo*	‘look at’
		ilo(i falaina)	‘search hair (i.e. for lice)’ (*falaina* ‘hair’)
		ilo(nunu)	‘a reflecting pool or mirror’ (*nunu* ‘shadow, image, picture’)
SES:	‘Are’are	*iro*	‘look for, collect’
		iro-iro	(N) ‘reflection, mirror’
SES:	Sa’a	*iro, iro-iro*	(VT) ‘look for, collect s.t.’
		irohi-	(VT) ‘clear the head of lice’
		iro-iro	(N) ‘a pool among rocks used as a mirror’
SES:	Arosi	*iro*	(VI) ‘look for, collect’
		irohi-	(VT) ‘look into, gaze into s.t., look at s.t.’
		(ha)irohi-	‘look for lice in the hair’ (*ha-* ‘verbal prefix’)
NCV:	Mota	*tiro*	(VI) ‘be clear’
		tiro(nin)	(N) ‘a little pool of water used as a mirror’ (*nin* ‘shadow, reflection’)
NCV:	Tamambo	*tiro*	‘look’
NCV:	Raga	*siroi*	(VT) ‘look steadfastly at s.t.’
NCV:	N. Efate	*ti-tiro*	(N) ‘mirror’
Mic:	Woleaian	*suro*	(VI) ‘look, watch, glance’
Fij:	Wayan	*tidro*	(VI) ‘look, peer, watch attentively’
		tidrovi-	(VT) ‘take a close look at s.t.’

PPn *tiro ‘look, observe’, *tiro-fi ‘gaze at s.t.’

Pn:	Niuean	*tio*	‘glance’
Pn:	Tongan	*sio*	(VI) ‘look, see’
		sio-fi	(VT) ‘keep one’s eyes fixed on s.t’
		sio-ʔi	‘peer at, look at in a critical or offensive way’
Pn:	W. Futunan	*jiro-a*	‘look carefully, search for’

Pn:	Pukapukan	*tilo*	‘gaze upon’
Pn:	Samoan	*tilo-tilo*	‘peep, peer, survey, look over’
		tilofi-a	‘be looked at, gazed at’
Pn:	Tikopia	*tiro-tiro*	‘look in pool as a mirror’
Pn:	Maori	*tiro, ti-tiro*	(VI) ‘look, look into, examine'

POc *kilop, *kilop-i- is reconstructable on the basis of two Polynesian witnesses with support from external witnesses in Western Malayo-Polynesian and Central Malayo-Polynesian.

PMP *kilep ‘glance, glimpse’ (ACD)

POc *kilop (VI) ‘glance’, *kilop-i- (VT) ‘glimpse s.t.’

Pn:	Tongan	*kilo*	(VI) ‘glance to one side, look out of the corner of the eyes’
		kilo-kilo	(VI) ‘glance from side to side, keep a sharp lookout’
		kilofi	(VT) ‘keep glancing at s.t.’
Pn:	Niuean	*kilo*	(VI) ‘turn the head, look around’
		kilo-kilo	(VI) ‘look around’

The following is the only reconstruction we have made for *seeing* verbs with both stative and active forms, based on evidence from Micronesian and Polynesian languages.

PMP *kilat ‘open the eyes wide’ (ACD)

POc *kila(t) (VSt) ‘be seen clearly, discerned, recognised’, (VAct) ‘see clearly, discern, recognise’

NCV:	Araki	*kila*	‘look, watch in a certain direction’ (*k* usually reflects POc *g)
Mic:	Chuukese	*k~~i~~ra*	(VT) ‘see, behold, find s.t.’
		k~~i~~ra-	(VSt) ‘be seen, found’ (in compounds only)
Mic:	Ponapean	*kila(ŋ)*	(VT) ‘see, discern, look at, observe, examine'
Mic:	Woleaian	*xa-xira*	(VT) ‘recognise it’ (*xa-* causative)
		xira	(VSt) ‘be clear, seen clearly, recognised’
Pn:	Tongan	*ki-kila*	(VI) ‘look with wide-open eyes’
Pn:	Rennellese	*kiga*	(VSt) ‘be clearly seen, in plain sight’

Polynesian reflexes of POc *(s,j)ila(k) ‘glance around’ sometimes refer to the mental attitudes attributed to someone glancing at something or somebody.

PMP *zilak ‘cross-eyed’ (ACD)2

POc *(j,s)ila(k) ‘look sideways, glance around’

SES:	Bauro	*sira-ia*	‘see’

PCP *jila, *ji-jila ‘look sideways’

Fij:	Rotuman	*jila*	‘(subj. eyes) squint, be crossed’

PPn *sila ‘glance, look sideways’ (POLLEX)

Pn:	Tongan	*hila*	(VI) ‘turn eyes away, glance’
		hila-ʔi	(VT) ‘glance at s.t., look at sideways’
		hile-hila	‘keep glancing’
Pn:	Niuean	*hela*	(VI) ‘to glance, look around furtively’ (*i > e irreg.)
		he-hela	(VI) ‘look, appear’
		hela-hela	(VI) ‘glance around’
Pn:	Rennellese	*siga*	‘look at, glance’

Pn:	Pukapukan	*yi-yila*	‘eyes opened wide’
Pn:	Samoan	*sila-sila*	‘see, watch’
		si-sila	‘stare, look steadily at’
		sila-fia	‘know’
Pn:	Maori	*hi-hira*	‘shy, suspicious’
Pn:	Tahitian	*hira*	‘bashfulness’
Pn:	Hawaiian	*hila-hila*	‘bashful, shameful, ashamed’

In a number of Oceanic languages patterns of polysemy indicate a close association between *seeing* and *knowing*. That vision is our primary source of objective data about the world is supported by child-language studies and by cross-linguistic studies of evidentials (Sweetser 1990:39). In Oceanic languages a *seeing* verb always refers to sight alone, never including other senses. The association between seeing and knowing is illustrated in reflexes of POc *kita and *re(k,q)i above and of POc *qilo below. The latter has been tentatively reconstructed as ‘be aware of, discern, see’. The most detailed evidence is from the Polynesian glosses, and this indicates that ‘know, be aware, recognise, notice’ is the core meaning, with ‘see’ as an extension.

POc *qilo ‘be aware of, discern, see’

MM:	Nakanai	*hilo*	‘to see’ (cf. *liho* ‘to see, look at’)
		hilo(tavu)	‘to think of, keep in mind’ (*tavu* ‘have contact with’)
NCV:	Mota	*ilo*	‘see’
NCV:	Raga	*ilo*	‘know, perceive’
NCV:	Tamambo	(*h*)*ilo*	(VI) ‘look while facing’ (*h* irregular)
Fij:	Wayan	*ilo-ilo*	(VI) ‘look, observe, watch’; (N) ‘glass (generic); mirror, looking glass’
		ilo-vi-	(VT) ‘notice, observe s.t.’
Fij:	Bauan	*ilo*	‘look at, as a reflection in water or in a mirror’

PPn *qilo ‘perceive, be aware of’ (POLLEX: ‘to know’)

Pn:	Tongan	*ʔilo*	(VT) ‘to see, espy, catch sight of, notice, perceive; find out, discover; be conscious or aware of; know, recognise’
		ʔilo-ŋa	(VSt) ‘show, show up, be seen, shown, recognised, known; conspicuous’
Pn:	Rennellese	*ʔigo-ŋa*	(N) ‘symbol’ (*igo-igo* ‘look, esp. at a reflection’)
Pn:	Samoan	*ilo*	‘perceive, be aware of’
Pn:	Tikopia	*iro*	‘take care of self or others’

Gedaged (NNG*) il* (V) ‘look at, behold, discern, perceive’; (N) ‘sight, view’ could reflect either *qilo or *kita, and we have no way of choosing between them.

We have considered the possibility that *qilo derives from POc *liqos (see above) by metathesis. Certainly the set of cognates supporting *qilo is in near-complementary distribution (according to subgroups) with those supporting *liqos. One could argue that Nakanai *hilo* is an independent development from the metathesis in Remote Oceanic languages. However, the semantic range of reflexes of *qilo, especially in Polynesian, appears to differ from *liqos reflexes.

A number of additional Polynesian forms are derived from POc *qilo. These forms are cited because they throw further light on the semantic range of *qilo.

PPn *qiloqilo (Adj*)* ‘be wise, aware’

Pn:	Tongan	*ʔilo-ʔilo*	(Adj) ‘be discerning, perspicacious, shrewd’, (VT) ‘know to some extent, have an idea of’
Pn:	Niuean	*ilo-ilo*	(Adj) ‘wise, clever’, (VI) ‘be clever’
Pn:	Rennellese	*ʔigo-ʔigo*	‘to know, understand, be aware of; be wise’
Pn:	Samoan	*ilo-ilo*	(VT) ‘investigate, examine s.t.’
Pn:	Tikopia	*iro-iro*	‘watching out, alerted, warned’

PPn *faka-qiloqilo ‘make s.o. wise’

Pn:	Tongan	*faka-ʔiloʔilo*	‘to teach, train, accustom’
Pn:	Niuean	*faka-iloilo*	(VT) ‘to be wise’
Pn:	Tikopia	*faka-iroiro*	‘to warn’

PPn *qilo-a (VI) ‘to know, be aware’, (VT) ‘know s.t.’

Pn:	Tongan	*ʔilo-a*	(VT, VSt) ‘be known, well-known, visible, within sight’
Pn:	Niuean	*ilo-a*	(VI) ‘to know’
Pn:	Pukapukan	*ilo-a*	‘know, understand’
Pn:	Samoan	*ilo-a*	(VT) ‘see, spot, notice, recognise, know, be aware of s.t.’
Pn:	Tikopia	*iro-a*	(V) ‘to know’, (VT) ‘know s.t.’
Pn:	W. Futunan	*iro-a*	‘to know’

PPn *faka-qilo-a (VT) ‘make s.t. known’

Pn:	Tongan	*faka-ʔilo*	(VT) ‘make s.t. known, report s.t.’
Pn:	Niuean	*faka-ilo-a*	(VT) ‘inform, make s.t. known’
Pn:	Samoan	*faʔa-ilo*	(VI) ‘signal to s.o.’ (takes oblique object)
		faʔa-ilo-a	(VT) ‘show, make s.t. known, advertise’
Pn:	Tikopia	*faka-iro*	‘inform beforehand’

PPn *faka-qilo-ŋa (N) ‘mark, sign, signal’

Pn:	Tongan	*faka-ʔilo-ŋa*	(VI) ‘make signs, signal’, (VT) mark s.t., make a mark’, (N) ‘sign, signal, mark’
Pn:	Samoan	*faʔa-ilo-ŋa*	(N) ‘mark’

The following cognate sets support reconstruction of another ‘see’ verb to PWOc level.

PWOc *nasi (VI) ‘look’, *nasi- (VT) ‘look at, see’

NNG:	Gedaged	*nasi*	(VT) ‘see, look at, behold, perceive; to experience, undergo’
MM:	Ramoaaina	*nai*	(VI) ‘look’, (VT) ‘look at, see’
MM:	Patpatar	*nas*	(VT) ‘see, look at’
		na-nās	‘open eyes, look for, search, gaze about’
MM:	Tabar	*nasi*	‘look for’
MM:	Siar	*nos*	‘look for’

Although POc verbs like *kita ‘see’, *kita-i ‘see s.t.’ and POc *liqo(s) (VI) ‘see, look’, *liqos-i- (VT) ‘see s.t., look at s.t.’ have both a transitive and intransitive form, they tend to occur in utterances with a specific object. *liqos-i- or one of the other reconstructed ‘look’ forms can be used to signal that the act is intentional or more tightly focused.

It is rare for a *seeing* verb to be able to take either experiencer or stimulus as subject. We have reconstructed a single verb, POc *kila(t) (VSt) 'be seen clearly, discerned, recognised', (VAct) 'see clearly, discern, recognise', where reflexes in Micronesia and Polynesia show that the same verb may carry either interpretation. Elsewhere, we have located examples where a *seeing* verb is used intransitively as a stative verb with source as subject only in the Tongan and E Futunan reflexes of *kita with meaning 'appear, come into view'. Evans (2003:68) concurs with respect to their rarity, but remains open as to whether intransitive *kita was (in Evans' terms) Actor or Undergoer subject, or perhaps either. The balance of the evidence favours Actor subject only.

2.2 Verbs of hearing

All languages in our sample have a transitive verb with 'hear s.t.' as one of its senses or its only sense, though in a number of Eastern Oceanic languages this verb may be extended to perceiving by non-visual senses. POc *roŋoR- 'hear s.t.' is generally reconstructed. However, there are certain problems associated with the formal reconstruction, to be discussed below.

A substantial number of reflexes of *roŋoR-, distributed across different high-order subgroups, carry the meaning 'listen (to s.t. or s.o.)', and it is likely that this sense was part of its semantic range in POc. An intransitive form, POc *roŋoR 'hear', is also reconstructable. In just a few languages this form is reflected as a stative verb, 'be heard', with the sound or its source as subject. We have located reflexes with the meaning '[be] heard' only in Gela, the Fijian languages and Tongan. This limited distribution suggests that the stative use has been developed independently in Gela and the Central Pacific languages. (Evans (1993) points out that in Philippine languages cognates show a similar uneven pattern of polysemy.)

It is likely that when Proto Oceanic speakers wished to comment on the nature of a sound they used the source as subject of a sound-specific verb, as *the drum is sounding*, *the leaves are rustling, their voices were audible* etc. Perception is implied, but the lexemes are not derived from verbs of perception. Oceanic speakers have a considerable vocabulary for the names of particular sounds, typically using them as both noun and verb. One of the more common ones is a reflex of POc *taŋis, an intransitive verb usually translated by 'cry', used to describe any sound characteristic of its source, as a cock crowing, dog howling, drum beating etc. In Toqabaqita, for instance, one could say *suʔari e aŋi ka falufalu* 'the drum is loud' (*suʔari* 'drum', *aŋi* 'to cry', *falufalu* 'sound loudly, of a drum' (lit. 'The drum is making a sound and it is loud'.) Samoan uses a term for 'voice', *leo* with verbal meaning 'sound' as in *e leo taʔe* 'it sounds cracked' *(taʔe* 'cracked'). The following is a random sample of sound terms: Tolai *tin* 'sound, as a coconut falling to the ground', *del* 'sound as the beating of a drum', *luluga* (N,VI) 'sound, as wind or rain'; Toqabaqita *ākwaʔa* 'make a slapping sound, as of a flat object', *ŋalu* 'of the sound of talking, be audible', *kutakuta* 'make a relatively loud, vibrating, pulsating sound'; Niuean *pakō* 'make a knocking sound', *kalī* 'make a rustling sound', *pakē* 'make a light crackling sound'.

Listening to something is sometimes given an extended cognitive meaning. In a number of languages (Gedaged, Nakanai, Nehan, Sursurunga, Sa'a), 'hear/listen' has been extended

to 'understand'.[6] A different extension of meaning is noted in many Southeast Solomonic and Central Pacific witnesses, where the meaning 'obey, take notice of s.o.' is present alongside 'hear, listen'. In Lakon (NCV) *ruŋ* means 'hear, feel', but also 'obey' and 'know' (Alexandre François pers. comm.). In Central Pacific languages this sense is usually associated with reflexes of *paka-roŋoR, which contains the intensifying prefix *paka-.

Certain difficulties arise in the reconstruction of the POc form(s) for 'hear'. We concur with Blust (ACD) who proposes POc *roŋoR, with initial *r, as the regular continuation of PMP *deŋeR. This is supported by non-Oceanic, Eastern Oceanic and by Schouten evidence. *loŋoR was a Western Oceanic variant, reflected in all WOc languages in which reflexes occur, except in the Schouten languages (Wogeo, Kaiep, Kairuru, Ali, Sissano and Sera).

POc *roŋoR- 'hear s.t., listen to s.t.'

Adm:	Lou	*roŋ*	'hear'
Adm:	Mbunai	*roŋ*	'hear'
Adm:	Seimat	*hoŋ*	(VT) 'hear, notice, become aware of, perceive'
NNG:	Kaiep	*(a)roŋ*	'hear'
SES:	Bugotu	*roŋo*	(VI) 'hear'
		va-roŋo	(VI) 'hear, listen to, obey'
		roŋo-vi	(VT) 'hear s.t., listen to s.t.'
SES:	Gela	*roŋo*	(VI) 'hear or be heard; listen, feel, obey; enquire about'
		roŋo-vi	(VT) 'hear s.t., listen to s.t.'
SES:	Lau	*roŋo-a*	'hear, listen to; perceive, smell'
SES:	Kwaio	*loŋo-a*	'listen, hear'
SES:	Sa'a	*roŋo*	'hear, listen, hear tidings of, understand'
SES:	Arosi	*roŋo*	(VT) 'hear, listen, obey'
SES:	Bauro	*roŋo-a*	'hear'
TM:	Teanu	*leŋi*	'hear'
TM:	Lovono	*laŋe*	'hear'
NCV:	Mota	*roŋo*	'apprehend by senses, hear, smell, taste, feel by touch'
NCV:	Lakon	*ruŋ*	'hear, feel; obey, know'
NCV:	Raga	*roŋo*	'hear, feel, apprehend by senses'
NCV:	Tamambo	*roŋo*	'hear, feel s.t.'
NCV:	Paamese	*loŋe*	(VT) 'hear, listen to; feel; pay attention to'
SV:	Kwamera	*reŋi-*	'feel, hear, smell, taste, perceive'
Mic:	Woleaian	*roŋo-roŋo*	'hear, listen to'
Mic:	Carolinian	*roŋ*	'to hear, listen, obey'

PCP *roŋo 'hear, be heard', *vaka-roŋo 'listen, heed, obey'

Fij:	Wayan	*roŋo*	(VSt) 'be heard, sound, be audible'
		roŋo-	(VT) 'hear s.t., listen to s.t.'
		vā-roŋo	(VI) 'listen, obey, heed'

[6] There are examples in Oceanic languages where 'understand' is also an extension of 'see', e.g. Kwamera *ata* 'see, look, regard, understand' and Niuean *kite* 'see, learn, understand, know', both reflexes of POc *kita 'see'.

Fij:	Bauan	*roŋo*	(VI) ‘hear, be heard’
		vaka-roroŋo	(VI) ‘listen, hear, obey’
		roŋo-ða	(VT) ‘hear s.t.’
Pn:	Tongan	*oŋo*	(VSt) ‘sound, be heard, be perceived’
		oŋo-ʔi	(VT) ‘hear s.t.; perceive, feel – pain, pleasure, taste, smell etc.’
Pn:	Samoan	*loŋo*	(VI) ‘perceive, by hearing or some other sense not sight’
		faʔa-loŋo	‘hear, listen; pay attention; obey’
Pn:	Rennellese	*goŋo*	(VI) ‘hear, listen, feel, taste’
		haka-goŋo	(VI) ‘hear, listen, obey, feel, taste’
Pn:	Maori	*roŋo*	(VT) ‘apprehend by the senses except sight; obey’

Variants with final *-n occur both with *r- initial and *l- initial forms, as in the following cognate set. Kove and Malalamai forms reflect either *roŋoR or *loŋoR, but given that all surrounding languages reflect *l- it would be odd if they did not.

POc *(r,l)oŋon ‘hear’

NNG:	Kove	*(i)loŋon-i*	‘hear’
NNG:	Malalamai	*(i)loŋon*	‘hear’
SES:	Talise	*roŋon-i-a*	‘hear’
SES:	Longgu	*roŋon-i-a*	‘hear (it)’
NCV:	Lepaxsivir	*roŋon-i*	‘hear’ (Tryon 1976:456–458)
Fij:	Wayan	*vaka-roŋon-i-*	(VT) ‘make s.t. known, cause s.t. to be heard’
Pn:	Tongan	*oŋon-a*	(VT) ‘hear, perceive, feel, be aware of (pain, pleasure, taste, smell), be aware or conscious of s.t., feel for, sympathise with’
Pn:	Niuean	*loŋon-a*	(VSt) ‘be heard’
		(fe)loŋon-āki	(VI) ‘hear each other’
Pn:	Samoan	*laŋon-a*[7]	(VSt) ‘be heard’
		faʔa-loŋo	(VT) ‘listen to s.t., hear s.t.’
Pn:	Tikopia	*roŋo*	(VI) ‘have bodily sensation, esp. hear, listen, feel’
		raŋon-a	(VSt) ‘be heard, felt’
Pn:	Maori	*roŋo*	(VT) ‘apprehend by the senses except sight’
		raŋon-a	(VSt) ‘be heard’
		whaka-roŋo	‘cause to hear, listen, attend to, obey’

Ross has hypothesised that the change in the final *-R to *-n and in the initial *r- to *l- was dissimilatory: to avoid two different trills (*R and *r) in the same very common word (Malcolm Ross, pers. comm.).

The following cognate set brings together some of the languages that reflect initial *l-:

POc *loŋoR (VI) ‘hear’, *loŋoR-i- (VT) ‘hear/listen to s.t.’

NNG:	Manam	*loŋor-*	‘hear s.o./s.t.’
		loŋor-i	‘obey, listen’
NNG:	Gedaged	*(i)loŋ*	‘know, have knowledge of, be aware of, hear, learn, perceive, understand’

[7] In Samoan and other Nuclear Polynesian languages PPn *roŋo-na ‘be heard’ shows an irregular change *o > *a* in the first vowel.

PT:	Bwaidoga	*nogala*	‘hear, listen to’
PT:	Gumawana	*nowo*	‘perceive s.t.; hear, listen, smell, sense s.t.’
PT:	Tawala	*nonola*	‘hear, smell’ (exp. *nogola*)
PT:	Sudest	*loŋwe*	‘hear’
PT:	Kiriwinan	*lagi*	‘hear, listen’
MM:	Bali	*loŋor-i*	‘hear’
MM:	Nakanai	*lolo*	‘hear, understand, know’
MM:	Meramera	*loŋ(e)*	‘hear’
MM:	Tiang	*loŋo-i*	‘hear’
MM:	Nalik	*laŋar*	‘hear’
MM:	Sursurunga	*a-loŋr-a*	‘hear; listen and understand’
MM:	Konomala	*luŋu-i*	‘hear’
MM:	Tolai	*va-loŋor*	‘hear’
MM:	Label	*loŋor*	‘hear’
MM:	Ramoaaina	*loŋoro-i*	(VI,VT) ‘hear, listen, heed, obey’
MM:	Siar	*loŋra-i*	‘hear’
MM:	Nehan	*loŋoro*	‘hear, understand’

The next set, although theoretically supporting a putative POc *noŋo(-noŋo), may simply reflect a number of parallel changes to *loŋoR or *roŋoR in which different languages independently assimilated initial *l- or *r- to the medial nasal.

Adm:	Mussau	*noŋo-noŋo*	‘hear’
NNG:	Matukar	*noŋ*	‘hear’
SJ:	Kayupulau	*nono*	‘hear’
SJ:	Ormu	*nono*	‘hear’
MM:	Banoni	*noŋono*	‘hear’
Pn:	Niuean	*fa-noŋo-noŋo*	(VI) ‘to listen’
Pn:	Tongan	*fa-noŋo-a*	(VI,VT) ‘to listen, hear, hear about’

2.3 Verbs of smelling

Verbs of smelling in Oceanic languages typically have an intransitive use, in which the source of the smell is the subject, and a transitive use in which the perceiver is subject and the source is direct object. Some examples follow.

Table 2: Some verbs of smelling that take both actor and source as subject

	Intransitive		Transitive	
PT: Saliba	*pane*	emit a smell	*pane-*	smell s.t.
MM: Minigir	*saŋina*	to stink	*saŋine-*	smell s.t.
MM: Tolai	*aŋina*	s.t. smell (sweet etc.)	*aŋine*	smell s.t.
SES: Gela	*aŋi*	emit strong smell	*aŋi-hi*	smell s.t.
SES: Kwaio	*moko*	to smell, stink	*moko-fi*	smell s.t.
Fij: Wayan Fijian	*garu* + modif.	to smell of s.t.	*garu-ti*	smell s.t.
Pn: Tongan	*namu*	emit smell	*nāmu-ʔi*	smell s.t.

The reconstruction of POc *s[a,o]ŋin (VSt) ‘emit a smell’, *s[a,o]ŋin- (VT) ‘smell s.t.’ is well supported if we accept that this form underwent sporadic changes in both the initial and final vowel. From the range of glosses exhibited by reflexes, it seems likely that *s[a,o]ŋin had both actor subject and source subject interpretations.

POc *s[a,o]ŋin (VSt) ‘emit a smell’, *s[a,o]ŋin- (VT) ‘smell s.t.’,

MM:	Bola	*(bu)roŋi*	‘(s.o.) smell s.t.’
MM:	Bola-Harua	*(bo)roŋi*	‘(s.o.) smell s.t.’
MM:	Sursurunga	*saŋin*	(VSt) ‘stink’
MM:	Patpatar	*saŋin*	(VSt) ‘produce an odour’
MM:	Minigir	*saŋine*	(VT) ‘smell s.t.’
		saŋina	(VSt) ‘stink’
MM:	Tolai	*aŋine*	(VT) ‘smell s.t., sniff at s.t.’; (Adj) ‘stinking, smelly’
		aŋina	(N) ‘smell, odour’; (VSt) ‘to smell ... (putrid, sweet, etc.)’ (In compounds usually *aŋ* only, e.g *aŋ na boroi* ‘to smell of pigs’)
MM:	Ramoaaina	*aŋina*	‘stink’

PMic *saŋu ‘smell s.t.’ (Bender et al. 2003) (*i > *u irreg.)

Mic:	Marshallese	*(ya-)teŋ*^{w}	‘smell s.t.’
Mic:	Puluwatese	*təŋ~~i~~ ~~i~~-(w)*	‘smell, sniff s.t.’
Mic:	Satawalese	*ttəŋ*	‘smell’
Mic:	Kosraean	*(m~~i~~-)sʌŋsʌŋ*	‘smelly, stinking of urine’ *(m~~i~~-* ‘urine’*)*

PPn *soŋi (VT) ‘smell s.t., sniff s.t., greet s.o. by pressing nose to face or limb and sniffing’

Pn:	Tongan	*hoŋi*	(VT) ‘sniff s.t. up, as in smelling salts’
Pn:	Niuean	*hoŋi*	(VT) ‘smell s.t., sniff s.t.’
		honoŋi	(VT) ‘(of humans only, active perception) smell s.t.’
Pn:	E Futunan	*soŋi*	‘touch noses’
Pn:	Pukapukan	*yoŋi*	‘smell, sniff, kiss’
Pn:	Rennellese	*soŋi*	‘press noses, kiss’
Pn:	Samoan	*soŋi*	(VT) ‘smell, scent s.t., smell s.o.’s cheek or hand, a method of kissing’
Pn:	Tikopia	*soŋi*	(VT) ‘smell s.t., sniff s.t., greet by nose pressure, accompanied by gentle sniffing, to face, wrist, knee, etc.’, (VI) ‘sniff’
Pn:	Tahitian	*hoʔi*	‘smell; touch noses’
Pn:	Maori	*hoŋi*	(VT) ‘smell s.t., sniff s.t., touch noses in greeting’
Pn:	Hawaiian	*honi*	(VT) ‘smell s.t., sniff s.t., touch noses in greeting’
cf. also:			
Adm:	Drehet	*hunu-huŋ*	‘smell’
MM:	Lavongai	*sain*	‘smell s.t.’
NCV:	Litzlitz	*suŋ*	‘smell’
NCV:	Timbembe	*suŋu-suŋu*	‘smell’

POc *quruŋ reflexes support its reconstruction as both a stative verb ‘emit a smell’, and a transitive verb *quruŋ-i (VT) ‘to smell s.t.’.

POc *quruŋ (VSt) ‘emit a smell’, *quruŋ-i (VT) ‘to smell s.t.’

NNG:	Bilbil	*ruŋ(ade)*	‘smell (s.t.)’
NNG:	Gedaged	*ìluŋ(an)*	‘smell (s.t.)’
PT:	Molima	*ulu(ma)*	(VSt) ‘to stink’
SES:	Gela	*uru*	(N) ‘a smell, good or bad’; (VSt) ‘to emit a smell’
		uru(mi)	(VT) ‘smell s.t.’
		uru(dika)	(VSt) ‘to stink’ (*dika* ‘bad’)
SES:	Lengo	*ur-uru*	(VSt) ‘emit a smell’
		uruŋi-a	(VT) ‘smell s.t.’
cf. also:			
MM:	Ramoaaina	*luŋi*	‘smell s.t.’

A separate term for the action of deliberately sniffing at something, POc *asok[8] ‘to sniff, kiss’ *asok-i ‘sniff or kiss s.t.’, is reconstructable. This term continues a PAn etymon, *Sajek. In a number of SES languages this action is represented by a compound, e.g. Kwaio *moko-toʔona (moko* ‘smell’*, toʔona* ‘put to the test, try*’),* Toqabaqita *makwa-toʔona* (VT) ‘check the smell of s.t.’.

PAn *Sajek (N) ‘smell’; (VT) ‘to smell (s.t.)’ (ACD)

PMP *hajek ‘smell, sniff, kiss’[9]

POc *asok (VI) ‘to sniff, kiss’ *asok-i (VT) ‘sniff or kiss s.t.’ (ACD)

Adm:	Wuvulu	*ato*	‘to sniff, smell’
Adm:	Seimat	*aso-i*	‘to sniff, smell’
PT:	Molima	*yaso*	‘to smell s.t.’
MM:	Nakanai	*aso-a*	(VT) ‘to sniff, smell s.t.’
MM:	Lamasong	*so*	‘smell s.t.’
MM:	Selau	*soka*	‘smell s.t.’ (metathesis)
Mic:	Kiribati	*aro(boi)*	(N) ‘smell, scent, the sense of smell’
		arok-i	(VT) ‘to smell or scent an odour’
Fij:	Rotuman	*aso*	‘to kiss by sniffing the face’
cf. also:			
SES:	Lau	*gasu*	(VSt) ‘to smell bad, stink’
Mic:	Carolinian	*uwas*	(N) ‘aroma or smell in the air, good or bad’

The forms listed below point to a POc verb *bona(s) (VI) ‘to smell, stink’, *bonas-i- (VT), either ‘(s.o.) smell (s.t.)’ or ‘(s.t.) smell of (s.t.)’. This appears to be related to POc *bo[-], *boe- (N) ‘odour, scent’, *baw-an, *bo-an (N) ‘odour, scent’ discussed below, but it is not derived by any known derivational process and may simply be a matter of chance resemblance.

---

[8] Note the contrast of meaning with POc *pVŋu (VI) ‘blow nose, sniff, snort’.

[9] Numfor, a South Halmahera-West New Guinea language, has a reflex of PMP *hajek: *yas* ‘native manner of kissing by smelling the face’. This meaning is mirrored in an Oceanic cognate of *hajek only in Rotuman, but illustrated also in Motu *harahu-a* ‘to smell, kiss’, in the ’Are’are and Sa’a term *nono* ‘to kiss, place the face against, sniff’, in Mota *pupupun* or *punpun* ‘to snuff [sic] in the native way of kissing’ and in a number of reflexes of PPn *soŋi ‘smell s.t, sniff s.t., greet s.o. by pressing nose to face or limb and sniffing’.

POc *bona(s) (VI) ‘to smell, stink’; *bonas-i- (VT) either ‘smell (s.t.)’ or ‘(s.t.) smell of (s.t.)’

NNG:	Maenge	*bona*	(N) ‘unpleasant smells’
PT:	Motu	*bona*	(N) ‘smell, scent’
		bona-ia	(VT) ‘to smell’
PT:	Balawaia	*bona*	(N) ‘smell, odour’
		bona-ia	(VT) ‘to smell’
NCV:	Mota	*puna*	(VI) ‘to smell, stink’ (*punai* (N) ‘smell, scent’)
		pun-pun	‘to snuff in the native way of kissing’
NCV:	Kiai	*pona-ponasia*	(VT) ‘smell s.t.’
NCal:	Nêlêmwa	*bo*	(VSt) (s.t.) ‘smell’
Mic:	Woleaian	*bō*	(N,VSt) ‘smell, stink’
		bō (maṣ)	‘to stink, smell bad’ (*mash* ‘be rotten, spoiled’)
		bō (ŋas)	‘be fragrant, sweet-smelling’ (*ŋas* ‘good, nice’)
		bō (lap)	‘to stink of armpit smell’ (*lap* ‘be big, huge’)
Mic:	Carolinian	*b^{w}ō*	(N) ‘smell, odour, aroma’
		b^{w}ō (maṣ)	‘stink, smell rotten’
		b^{w}ō (ŋas)	‘be fragrant, sweet-smelling’
		b^{w}ō (pa)	‘smell of shit etc.’ (*pā* ‘faeces’)
Fij:	Bauan	*bona*	(VI) ‘stink because rotten; (N) stinking rottenness; a stench’
		bona-ða	(VI) ‘to stink of s.t.’

Lynch (2001) reconstructs the set below. It resembles POc *bona(s) but this may well be a chance resemblance, as PSV *-e- does not regularly reflect POc *-o-.

PSV *a-b(i)eni (VI) ‘emit an odour’ (Lynch 2001)

SV:	Sie	*empen*	(VI) ‘emit an odour’
SV:	Lenakel	*əpien*	(VI) ‘emit an odour’
SV:	Kwamera	*apein*	(VI) ‘emit an odour’
SV:	Anejom	*epeñ*	(VI) ‘emit an odour’
		nepñ(ami)	(VI) ‘smell of urine’
		epeñ(wañ)	(VI) ‘have musty smell’

Tryon (1976) lists a number of NCV languages which denote ‘smell s.t.’ by compounding reflexes of *roŋoR ‘hear’ with reflexes of *bona(s), to reflect PNCV *roŋo-bona, e.g. Raga *roŋ-bunina*, Lametin *roŋ-bonai.*

Our starting point for the cognate set below is a pair of PMP forms, *bahu (V) ‘smell bad’, reconstructed by Dempwolff, and *bahu-an (N) ‘odour, stench’, reconstructed by Blust (ACD) as a suffixed form of Dempwolff’s *bahu. Dempwolff glossed *bahu as a noun, but it seems likely that it was used as a verb, since *bahu-an includes the nominaliser *-an. Blust *reconstructs* *bahu-an as Proto Western Malayo-Polynesian, but the Oceanic forms listed below show that it occurred in PMP.[10]

Most Malayo-Polynesian languages have lost *h, with the result that *bahu and *bahu-an respectively became *baw and *baw-an. By regular sound change *baw probably became POc *bo, but *baw-an may have survived in this form in POc alongside

[10] Thanks are particularly due to Malcolm Ross for making a detailed analysis of these two forms. The following paragraphs are based substantially on his comments.

*bo-an (see discussion of Manam b^{w}au below). POc *bo is a phonotactically rare phenomenon, a monosyllabic lexical root. There has long been a tendency in Austronesian languages to make lexical forms disyllabic, and so *bo occurs with a number of extensions, some of which we cannot fully explain. Some of these extensions are disyllabic roots in their own right: see Bola *bu-roŋi*, Bola (Harua) *bo-roŋi,* both '(s.o.) smell s.t.' under POc *s[a,o]ŋin 'emit a smell' above, and the Woleaian and Carolinian examples below. In Central Pacific and Micronesian languages a monosyllabic lexical root becomes bimoraic, i.e. its vowel is long.

The POc noun *bo[-] is shown below with a bracketed hyphen, as a number of its reflexes are monovalent nouns. Monovalent reflexes in Ponapean and Puluwatese suggest that the POc monovalent form was sometimes *boe-, and it is perhaps this form that is also reflected by Kiribati p^{w}oi (zero-valency noun) and Bauan *boi* (intransitive verb).

PMP *bahu (VI) 'smell bad' (Dempwolff, cited by Blust, ACD)

POc *bo[-], *boe- (N) 'odour, scent'; *bo (VI) 'have an odour, be smelly'

PNCV *b[o,u][-] (N) 'odour, scent'; *b[o,u] (VI) 'have an odour, be smelly'

NCV:	Port Sandwich	*mbo*	'to stink, rotten'
		na-mbo-	'smell, odour'
NCV:	Uripiv	*o-po*	'rotten',
NCV:	Big Nambas	*-pu*	'it is rotten'
NCV:	Neve'ei	*bo*	'rotten, stink'
NCV:	Nati	*mpu*	'rotten, stinking'
NCV:	Labo	*pu-o*	'to stink'
		nu-mbu-	(N) 'smell'
NCV:	Naman	*-bu*	'stink; rotten'
NCV:	Sa	*bo-*	(N) 'smell'
NCV:	Lonwolwol	*bo*	(VI) '(s.t.) smell'
NCV:	Paamese	*vō*	(VI) '(s.t.) smell'
NCV:	Nguna	*p^{w}o*	(VI) 'stink, smell bad'
SV:	Sye	*(e)mpu*	(VI) '(s.t.) smell'
SV:	Ura	*(i)bu*	(VI) '(s.t.) smell'
NCal:	Nêlêmwa	*bo*	(VI) '(s.t.) smell'
Mic:	Kiribati	*p^{w}oi*	(N) 'smell, odour'
Mic:	Ponapean	*p^{w}ō, p^{w}owɛ-*	(N) 'smell, odour'
Mic:	Chuukese	*p^{w}ō*	(N) 'smell, odour'
Mic:	Woleaian	*f^{w}ō*	(N, VI) 'smell, stink'
		f^{w}ō(mas)	'to stink, smell bad' (*mash* 'be rotten, spoiled')
		f^{w}ō(ŋas)	'be fragrant, sweet-smelling' (*ŋas* 'good, nice')
		f^{w}ō(lap)	'to stink of armpit smell' (*lap* 'be big, huge')
Mic:	Carolinian	*b^{w}ō*	(N) 'smell, odour, aroma'
		b^{w}ō(mas)	'stink, smell rotten'
		b^{w}ō(ŋas)	'be fragrant, sweet-smelling'
		b^{w}ō(pa)	'smell of shit etc.' (*pā* 'faeces')
Mic:	Puluwatese	*p^{w}o, p^{w}o~~i~~-*	(N) 'smell, odour'
Fij:	Bauan	*boi*	(VI) 'have an odour'
		boi-ða	(VI) '(s.t.) smell of'

Among the reflexes of POc *bo-an below Manam *b^{w}au* requires particular comment. At first sight it looks as if it reflects a POc *baw, i.e. a form in which earlier *-aw has not become POc *-o. It is rather more likely, however, that it reflects POc *baw-an, with regular loss of final *-n and consequent irregular loss of *-a, since earlier word-internal *-aw- did not always become POc *-o-. Gumawana *bowana* is the only form below which attests to the presence of POc final *-n. Note that the Gumawana and Gapapaiwa forms both serve as verbs as well as nouns.

PMP *bahu-an (N) 'odour, stench' (ACD: PWMP)

POc *baw-an, *bo-an (N) 'odour, scent'

NNG:	Manam	*b^{w}au*	(N) 'smell, odour'
PT:	Gumawana	*bowana*	(VI) 'stink, smell bad'; (N) 'bad odour'
PT:	Gapapaiwa	*boa*	(VI) 'rot, smell bad'

PNCV *boa (N) 'odour, scent'

NCV:	Tambotalo	*poa*	'smell'
NCV:	Nguna	*na-p^{w}oa*	'smell'

PPn *poa (N) 'fish odour'

Pn:	Tongan	*poa*	'yam with fishy smell'
		(namu)poa	'fish odour'
Pn:	Niuean	*poa*	'fish odour'
Pn:	Anuta	*po-poa*	'fishy smell'
Pn:	E Futuna	*po-poa*	'fish odour'
Pn:	Samoan	*poa-poā*	'fish odour'
Pn:	Sikaiana	*poa*	'fish odour'
Pn:	Tokelauan	*poa-poā*	'smelling of fish'
Pn:	Marquesan	*poa (ika)*	'chum, bait'
Pn:	Rarotongan	*poa*	'fishy (smell or taste); scales, rust'
Pn:	Maori	*poa*	'bait; allure by bait, entice'

Blust (1988) has reconstructed a family of PAn 'stench' words which all contain the phonemic sequence *qaŋe- (*qaŋeSit 'stench, musky odor of an animal', *qaŋeliC 'stench of burning substances', *qaŋeRiS 'stench of fish', *qaŋeRu 'stench of spoiled or souring organic matter' and *qaŋeseR 'stench of urine'). The only clear trace of these in Oceanic languages is in the Gela term: *aŋo* 'emit a sour smell, as of urine'.

PAn *qaŋeSeR 'stench of urine' (Blust 1988, ACD)

POc *(q)aŋo(R) (VSt)'smell, as of urine'

SES:	Gela	*aŋo*	(VSt) 'emit a sour smell, as of urine'
cf. also:			
SES:	Arosi	*waŋo*	(VI) 'smell (sweet or otherwise)'
		waŋo-ra	'to smell of blood'

Oceanic languages often have terms for the smell of urine and other body secretions, and terms for various other odours, good and bad. Milner's Samoan dictionary, for instance, lists *soŋo* (V) 'smell of urine etc.', *lalaʔoa* (V) 'smell of fish', *sauŋa* (N,V) 'smell of stale food etc.', *ʔalalā* (N,V) 'smell of meat or fish when cooked', *elo* (V,Adj) 'give an offensive smell of decomposing flesh'. However, few terms for specific odours have been

collected from other languages and we have been unable to make reconstructions other than the one above and the following:

PMP *seŋet ‘acrid, pungent, of odour’ (ACD)

POc *soŋo ‘[be] acrid, pungent, as smell of urine’

NNG:	Lukep Pono	*-yoŋo*	‘smell s.t.’
MM:	Label	*soŋ*	‘smell (s.t.)’
MM:	Tiang	*(mo)soŋ*	‘smell (s.t.)’
MM:	Notsi	*coŋo*	‘stink’
Mic:	Carolinian	*(b^wō)ttoŋo-toŋ*	‘smell sweaty, unclean, unwashed’ (*b^wō* ‘odour’*)*

PPn *soŋo ‘smell of urine’ (POLLEX)

Pn:	Niuean	*ho-hoŋo*	‘perceive an odour, smell s.t. (as from a distance)’
Pn:	Tongan	*ho-hoŋo*	‘smell of urine’
Pn:	E. Uvean	*ho-hoŋo*	‘smell of urine’
Pn:	Pukapukan	*yo-yoŋo*	‘smell of urine’
Pn:	Rennellese	*soŋo(aŋa)*	‘sex organs’
Pn:	Samoan	*soŋo*	‘(of urine, etc.) smell, stink’
Pn:	Nukuoro	*soŋo-soŋo*	‘genitals (male or female)’
Pn;	Tikopia	*soŋo*	‘female genitalia’
Pn:	Tokelauan	*so-soŋo*	‘smell of urine’
Pn:	Hawaiian	*ho-hono*	‘odour of perspiration’

PMP *maŋsit ‘vile smell’ (ACD)

POc *masi(t) ‘smell bad; [be] sour, acid, fermented’ (N) ‘bad smell’

NNG:	Manam	*masi*	‘smell of fish’
		masi-masi	(VI) ‘smell bad’
PT:	Ubir	*mas*	‘to smell’
SES:	Gela	*mahi*	‘body smell’
SES:	Arosi	*masi*	(N,V) ‘smell of stale fish or urine’
		masi(ŋaʔi)	(VSt) ‘smell stale, sour’ (*ŋaʔi* ‘verb suffix’)
Mic:	Carolinian	*m^was*	(VI) ‘stink’
Pn:	Tongan	*mahi*	‘sour to the taste’
Pn:	E Futunan	*masi*	‘acid, fermented, preserved by fermenting’
Pn:	Samoan	*masi*	‘fermented breadfruit’
Pn:	Maori	*mahi-mahi*	‘rotten, putrid’

Familiar smells may be lexicalised, either as a stative verb or noun, e.g. *(q)aŋo(R) ‘smell, as of urine’ (from PAn *qaŋeSeR ‘stench of urine’), *soŋo ‘[be] acrid, pungent, as smell of urine’ (from PMP *seŋet ‘acrid, pungent, of odor’), and *masi(t) ‘smell bad; [be] sour, acid, fermented’ (from PMP *maŋsit ‘vile smell’). Although these three POc reconstructions have here been given a verbal form, it is evident that in two cases the PAn or PMP antecedents are nouns and in the third, many of the lower level reflexes are also nouns. It is noteworthy that in all the cognate sets supporting verbal ‘smell’ reconstructions there are examples of the verb functioning also as a noun. This tendency has not been noted in any of the other sense-related verbs other than in the terms given to specific sounds. We have reconstructed one generic noun, POc *bo[-] ‘odour, scent’ which can also be used as a stative verb, meaning ‘have an odour’.

2.4 Verbs of tasting

Taste is the sense that informs us about what we are eating or drinking. Perception of taste is usually the outcome of an intentional act. POc *ñami- (VT) 'taste s.t., test the flavour of food' is well attested, with reflexes scattered across diverse subgroups. Reflexes of a partially reduplicated form, *ña-ñami, also occur in some languages as a stative verb, meaning 'be tasty, taste good'. The attribution of this sense to POc is somewhat strengthened by extra-Oceanic cognates. A fully reduplicated form, *ñami-ñami is also reflected in Ramoaaina (MM), Gela and Longgu (SES), Marshallese (Mic) and Rennellese (Pn), and this may have been an intransitive verb meaning 'to taste, do tasting'.

*ña-ñami has a PMP antecedent in the form of PMP *ñamñam 'taste, tasty' which Blust (1989) reconstructs on the basis of Tagalog *namnam* 'savor, taste; palatal sensation' and Selaru *nanam* 'sweet' together with Oceanic reflexes. POc *ñami appears to continue the PMP root *ñam with the addition of the transitive sufix *-i.[11]

Except for Tikopia, which has doublets *nami* 'taste' and *namu* 'odour, bad smell', the Central Pacific reflexes of *ñami show a vowel change *i > u. Polynesian reflexes tend to blur the distinction between taste and smell, both senses contributing to the assessment of quality of food. The shift from flavour to odour is complete in Pukapukan, Rennellese and Samoan.

PMP *ñamñam 'taste, tasty' (Blust 1989)

POc *ña-ñami (VI) '[be] tasty, taste good', *ñami- (VT) 'to taste s.t.'

Adm:	Nyindrou	*ñimi-ñem*	(VT) 'taste, test flavour of'
MM:	Patpatar	*nam-nami-en*	'be tasty, sweet'
		nami-en	(VT) 'taste s.t.'
MM:	Tolai	*namene*	(VI,VT) 'to taste'
MM:	Ramoaaina	*nam-nami-an*	'sweet'
SES:	Gela	*nami*	(VT) 'to taste s.t.; tasting'
		nai-nami	'to taste, tasting'
SES:	Bugotu	*ñami*	(VT) 'to nibble, bite, taste s.t.'
SES:	Ulawa	*name*	(VI) 'to taste'
		name-li	(VT) 'to taste s.t.'
SES:	Arosi	*nami*	(VI) 'to taste'
		nami-ri	(VT) 'taste, lick s.t.'
SES:	Longgu	*nami-*	(VT) 'taste s.t.'
		nami-nami	(VI) 'taste'
NCV:	Mota	*nam, nami-s*	'to taste, touch with the tongue'
NCV:	Mwotlap	*nɛm*	'taste s.t. with tongue, lick'
NCal	Nêlêmwa	*nām*	'sweet'

PMic *ñama 'taste' (also *ñaña 'taste, flavour') (Bender et al. 2003)

Mic:	Kiribati	*na-nama*	'to taste or test the flavour of s.t.'
Mic:	Woleaian	*nana*	'taste, try the taste of'
		na-ri	(VT) 'taste, try s.t.'
Mic:	Carolinian	*nanna*	(VI) 'to have a certain flavour or taste'

[11] There are many parallel cases of PMP disyllables of the form R-R (where R is a monosyllabic root) being continued in Oceanic as R-i (Blust 1977, Ross 1998:24–25).

Mic:	Marshallese	*nam-nam*	‘taste, smell, flavour’
Fij:	Bauan	*namu*	‘chew and swallow’

PPn *namu (V) ‘taste’, (N) ‘odour, flavour’; *namu-aʔa (VSt) ‘have a strong smell or flavour’

Pn:	Tongan	*namu*	(VSt) ‘emit a smell’ (only in compounds, e.g. *namu-hohoŋo* ‘to smell of urine’, *namu kakala* ‘be fragrant’, *namu-kuu* ‘to stink’, *namu-toto* ‘smell of blood’ etc.)
		na-namu	(VSt) ‘emit an odour, to smell’; (N) ‘odour, smell’
		nāmu-ʔi	(VT) ‘perceive the smell/taste of’
		namu-aʔa	(VSt) ‘have a strong or pungent smell’
Pn:	Niuean	*namu*	(N) ‘odour, flavour’
		namu-ā	(VSt) ‘smell of fish or the sea’
Pn:	Pukapukan	*namu*	(VSt, N) ‘smell s.t., emit an odour typical of s.t., e.g. *namu ika* ‘fishy smell’, *namu ānani* ‘sweet smell’
		na-namu	(VSt) ‘very smelly, putrid’
Pn:	Rennellese	*namu-aʔa*	(VSt) ‘to stink’
		na-namu	(VSt) ‘to smell good or bad’
		namu-namu	‘to inhale, sniff, as at a distance’
Pn:	E Futunan	*namu-kū*	‘bad odour, flavour’
Pn:	Samoan	*nāmu*	(VSt) ‘to smell of, have the odour of’
Pn:	Tikopia	*namu*	(N) ‘odour, smell (used of strong or unpleasant smells)’
		nami	‘taste in experimental way’
Pn:	Maori	*namu-namu-ā*	‘flavour imparted to food by contact with s.t.’
Pn:	Rapanui	*namu-namu*	‘to taste, chew’

There is a formal similarity between *ñami and the next two reconstructions, *tami and *mami, together with *(d,dr)ami ‘lick’ (cognate set not included here). They may share descent from a PAn monosyllabic root, *mis (Blust 1988).

PMP *tamiq, *tamis ‘taste, try’ (ACD)

POC *tami ‘taste, try’

MM:	Tolai	(*an-*)*tamai*	(VT) ‘to taste, of food’ (*an* ‘to eat’) (problematic vowels)
Pn:	Rennellese	*tami*	‘taste’
		tami-tami	‘taste a little, as to try’

It is possible that POc *mami ‘sweet’ has evolved by a different route from POc *mami ‘try by tasting’. Whereas the former is derived directly from PAn and PMP etyma, the latter may be the product of contamination between POc *mami ‘sweet’ and POc *ñami ‘to taste s.t.’.

PAn *ma-amis ‘sweet’ (Tsuchida 1976)

PMP *mamis ‘sweet’ (Dempwolff); *emis ‘sweet taste’ (ACD)

POc *mami ‘to try by tasting; sweet’

PT:	Motu	*mami-a*	(VT) ‘to feel, test’
PT:	Balawaia	*mami-*	(N) ‘taste’
SES:	Gela	*mami-a*	‘tasting good’
SES:	Sa’a	*mami*	(VI) ‘to taste'
SES:	Kwaio	*mami*	‘normal tasting, neither sweet nor sour’
		mami toʔona	‘try food, taste’ (*toʔo* ‘receive, catch’)

Mic:	Puluwatese	*məm*	‘sweet’
Mic:	Carolinian	*mam*	‘be sweet-tasting’
Mic:	Woleaian	*mami*	‘sweet’
Fij:	Wayan	*mami*	‘cooking banana, sweet-tasting’

POc *ñapi- (VT) ‘taste s.t.’ may have evolved from *ñami by the strengthening of medial *-m- to a prenasalised stop *b, with subsequent devoicing.

POc *ñapi- (VT) ‘taste s.t.’ (Blust 1998)

Adm:	Lou	*nap*	‘taste’
SES:	Gela	*na-napi*	(VT) ‘taste, lick s.t.’
SES:	Bugotu	*ñapi*	(VT) ‘to bite, taste s.t.’
SV:	Sye	*(at)ŋap*	‘taste’
SV:	Ura	*(ar)ŋap*	‘taste’

Some Oceanic languages lack a verb dedicated to the meaning of intentional tasting. Instead, they use a verb whose basic meaning is more general, such as reflexes of POc *topoŋ-i ‘try/attempt s.t.’ or ‘sample s.t.’, or terms for ‘nibble’ or ‘lick’. To limit such a verb to the tasting of food or drink a qualifier is added. Thus in Seimat one says *ŋa ani tohoŋi-wa* ‘I taste the food’ *(ani* ‘eat’*, tohoŋi* (VT) ‘try, attempt s.t.’) and in Tolai *an-tamai* (VT) ‘to taste, of food’ (*an* ‘eat’, *tamai* ‘taste, try’). Kwaio has *ʔana toʔona* (*ʔani* ‘eat’, *toʔona* ‘put to the test’) and *mea toʔona* (*mea* ‘tongue’, *toʔona* ‘put to the test’), both meaning ‘taste (food)’, while Toqabaqita has *ʔani-toʔona* ‘taste s.t. by eating it, try the taste of s.t.’ (*ʔania* ‘eat’, *toʔona* ‘test, check’) and *kuʔu-toʔona* ‘drink s.t. to see what it is like’ (*kuʔu* ‘drink’).

PMP *tepeŋ ‘try, test, experiment’ (Blust, pers. comm.)

POc *topoŋ (VI) ‘try’, *topoŋ-i (VT) ‘try, test, sample s.t.’

Adm:	Seimat	*tohoŋ-i*	(VT) ‘try, attempt’ (*ŋa ani tohoŋi-wa* ‘I taste the food’)
Adm:	Mussau	*tōtoŋa*	(VT) ‘taste’
NNG:	Gitua	*tovo*	‘try’
PT:	Gapapaiwa	*tovon*	‘feel, squeeze’
PT:	Motu	*(mami-a)toho*[12]	(VT) ‘to test and try’ (*mami-a* ‘to feel, test’)
MM:	Nakanai	*tovo*	‘measure, try out’
SES:	Gela	*tavoŋo*	‘grope, feel in the dark’
SES:	Sa'a	*ohoŋ-i*	‘to attempt, make trial of’
SES:	Arosi	*oho*	‘to contend’
		ohoŋ-i	‘try, test, tempt’
NCV:	Mota	*to-towo*	‘do for the first time’
Mic:	Ponapean	*soŋ*	‘taste s.t., attempt’
Fij:	Bauan	*tovo(le)*	(VI) ‘try, attempt, test’
		tovo(le)-a	(VT) ‘try s.t.’
Pn:	Samoan	*tofo*	‘test, sample’
cf. also:			
SES:	Gela	*tovo*	‘to ask for a wife’

[12] Compare Motu *daua-toho* ‘to feel a thing’ (*dau-* ‘stretch out the arm’, *toho* ‘try’)

A speaker of POc could pass judgement on the taste of something in the mouth by using an adjectival verb, with the source of the flavour as subject. The following verbs are specifically taste descriptors: *mami '[be] sweet, *ña-ñami '[be] tasty, taste good' and *maqasin '[be] salty' (Ross et al., 2003:68). A PPn verb *suqa (VI) 'have a certain taste or flavour' is reconstructable, occurring in the compound *suqa-malie 'taste good; sweet, delicious'. Another strategy is to use a compound of the form V + modifier, where V is a verb meaning 'eat' or 'drink', e.g. Wayan Fijian *kani vinā* 'be tasty (lit. 'eat well'), *kani ðakaðā* 'taste bad' ('eat bad').

While smelling terms are frequently nouns, taste terms are predominantly verbs. The only examples in our cognate sets where reflexes of a reconstructed verb are used as a noun are in Niuean and Tikopia reflexes of POc *ñami- (VT) 'to taste s.t.', and in each case the term has changed its meaning from taste to smell.

2.5 Verbs of perceiving by touch

The sense often labelled 'touch' has to do with perceiving pressure on the skin. Awareness of such pressure is expressed in English by the verb 'feel', e.g. 'I can feel the wind in my face'. More commonly the reference is to contact between skin, usually hand, and a solid object, resulting in awareness of some property of the latter's surface. 'Feel' is also used in English to denote awareness of a physiological or emotional condition, e.g. 'feel sick or frightened or responsible'. We will not be concerned here with the latter sense of 'feel'.

POc *si(g,k)il 'touch with the fingers' is our strongest candidate for a verb meaning 'perceive by touch'. Reflexes of POc *taŋo 'take hold of, grasp, touch with the hand' tend to carry the additional meaning of deliberately taking hold or grasping. In some languages reflexes of a PWOc term *sau 'reach out with hand, touch' may be combined with a verb meaning 'try' to express that meaning, as in the Motu and Nakanai examples below, but we cannot reconstruct a specific compound verb for PWOc meaning 'perceive by touch'. Other Oceanic languages use verbs that are either primarily verbs of manipulation (do s.t. by hand, grope, grasp, poke, stroke etc.) or of making contact in a physical sense, without involving awareness (be in contact, reach), although some may have had 'perceive by touch' as a secondary sense.

POc *si(g,k)il, *si(g,k)il-i- 'touch with the fingers'

MM:	Patpatar	*sigire*	(VT) 'touch, lay hands on to abuse'
MM:	Sursurunga	*sigil, siŋli*	(VT) 'touch'
SES:	Bugotu	*hīgili*	'touch s.t.'
SES:	Gela	*higili*	'touch s.t.'
		kisi, *gisi*	(VT) 'touch with finger, poke' (metathesis)
SES:	Sa'a	*siki*	'tap, touch with fingers'
		siki-hi	(VT) 'infect, carry infection (to s.o.)'
		siki-li	(VT) 'twang with the fingers'
SES:	Arosi	*sigi*	'tap with the finger'
		sigi-hi	(VT) 'infect with'
NCV:	Mwotlap	*hiɣ*	'poke, esp. with finger; point at'
NCV:	NE Ambae	*sikeli*	(VT) 'touch'
NCV:	Tamambo	*hisi*	(V) 'reach, touch s.t.' (metathesis)
NCV:	Namakir	*qih*	'touch, stroke' (metathesis)

NCV:	Nguna	*kisi*	‘touch with fingers’ (metathesis)
Mic:	Kiribati	*rī(ŋa)*	‘feel (s.t.), handle, touch’

POc *taŋo ‘take hold of, grasp, touch with the hand’

NNG:	Sio	*taŋo*	‘touch, place hand or fingers on’
SES:	Gela	*taŋo*	(VI) ‘do, touch, be in contact’(rarely used except in compounds)
		taŋo(li)	(VT) ‘hold, touch’
SES:	Bugotu	*taŋo(li)*	(VT) ‘to take, hold, handle, receive’
		taŋoli hadi	‘feel for a thing’ (*hadi* ‘go up’)

PNCV *taŋo-vi ‘touch, feel, grope’ (Clark 2009)

NCV:	Mota	*taŋo*	‘touch, feel with the hand’
NCV:	Paamese	*taŋo-taŋo*	‘place hands on s.t.’
NCV:	Sakao	*daŋ*	(VI) ‘to grope’
NCV:	Lonwolwol	*toŋve*	‘to touch’
Pn:	Pukapukan	*taŋo*	‘take hold of, grasp’
Pn:	Samoan	*taŋo*	(VT) ‘take hold of s.t., touch (and feel) with the hand, feel’
		taŋo-fia	‘be touched’
		taŋo-taŋo	‘lay hold of, touch and feel’
Pn:	W Futunan	*taŋo*	‘grope, feel for’

PWOc *sau ‘reach out with hand, touch’

PT:	Motu	*dau*	(VI) ‘stretch out the arm’, (VT) ‘touch, feel’(*dau-kunu* ‘to touch, when fingers touch an object’ (*kunu* ‘fill, be satisfied’), *dau-dae* ‘to stretch the arm up’, *dau-lata* ‘stretch out the arm for s.t. in front’ etc.)
		daua(toho)	‘feel a thing’ (*toho* ‘try’)
MM:	Nakanai	*sau*	‘place the hand’
		sau(lalai)	‘to feel tentatively (with hand)’ (*lalai* ‘to try’)

PNCV *tiqeli ‘touch, reach’ (Clark 2009)

NCV:	Paamese	*tokoli*	‘touch, feel with hands, reach, go as far as’
NCV:	Kiai	*tikeli-a*	‘touch’
NCV:	Lewo	*toli*	‘reach, arrive at, touch’
NCV:	NE Ambae	*sikeli*	‘touch, reach, arrive at’
SV:	Anejom	*etcai*	‘feel, touch’
cf. also:			
PT:	Kiriwina	*(mom)koli*	‘taste, sip’
MM:	Tolai	*tuk*	‘touch with hand or pointer’

3 Conclusions

For the five ‘basic’ senses we can reconstruct at least one POc transitive verb dedicated to a particular sense, namely: *kita- ‘see s.t.’, *roŋo- ‘hear s.t.’, *sa[a,o]ŋin- ‘smell s.t.’, *ñami- ‘taste s.t.’ and *si(g,k)il-i- ‘touch with the fingers’. Certain of these verbs were polysemous but each had a canonical use in which the grammatical subject is the experiencer of an act of perception and the direct object is the stimulus.

Three of the above reconstructions, *roŋo-, *sa[a,o]ŋin- and *ñami- are reconstructable with both 'sensing' and 'attending' senses, that is both with and without intention. Reflexes of *si(g,k)il-i are apparently typically used with an intentional force. In order to express the meaning 'listen' a number of languages (Southeast Solomonic, Fijian, Polynesian) add an intensifying prefix to *roŋo, providing some evidence for PEOc *paka-roŋo(R,n)- 'listen to s.t.'. This kind of semantic extension was probably not characteristic of *kita- 'see s.t.'. To denote intentional acts of visual perception POc speakers, like English speakers, could choose from a range of different transitive and intransitive verbs meaning, e.g. 'look (at s.t.)', 'glance', 'look intently or closely', 'peer (at s.t.)' and 'look for s.t.'. There are many more verbs denoting kinds of visual activities than there are verbs denoting kinds of hearing, smelling, tasting and sensing by touch, and many of the former involve intent.

With regard to meanings where the stimulus or source is subject, POc perception verbs vary in their ability to occur as stative verbs. With verbs of *seeing* and *hearing*, stimulus-subject verbs are very rare in daughter languages. A single reconstruction, POc *kila(t) (VSt) 'be seen clearly, discerned, recognised', (VAct) 'see clearly, discern, recognise', has been made. Languages tend instead to use verbs unrelated to the transitive forms to represent meanings like 'be visible/be seen/appear, be audible/be heard/sound'. When the focus is on the outcome of hearing, languages generally have a range of stative verbs comparable to 'be noisy, be loud'. We have collected a number of such terms but have made no reconstructions. Languages tend also to have many terms for specific sounds which can be used as stative verbs with source as subject. In such cases the act of perception is implied. A number of these verbs also act as nouns, a feature shared with those *smell* verbs that refer to specific odours.

In the case of *smelling* and *tasting*, however, stative verbs derived from actor-subject verbs are common in Oceanic languages and several such pairs have been reconstructed for POc, e.g. *s[a,o]ŋin (VSt) 'emit a smell' (alongside *s[a,o]ŋin- (VT) 'smell s.t.'), POc *quruŋ (VSt) 'emit a smell' (alongside *quruŋ-i (VT) 'to smell s.t.') and POc *ña-ñami (VI) '[be] tasty, taste good' (alongside *ñami- (VT) 'to taste s.t.'). For verbs of *smelling* and *tasting* it is also possible to reconstruct stative verbs that refer to qualities specific to one sense, as POc *soŋo 'be acrid, pungent', POc *masi(t) 'smell bad; [be] sour, acid, fermented', *mami '[be] sweet, *ña-ñami '[be] tasty, taste good' and **maqasin* '[be] salty', although only *ña-ñami, is derived from an experiencer-subject verb. The others can be attached to appropriate nouns without the need for a verb of sensing.

The variations in the linguistic expression of the different senses that we find in Oceanic languages are grounded, at least in part, in the nature of the senses themselves. Each human sense operates under certain conditions that influence the way it is expressed. *See* and *hear* have a degree of commonality in that the experiencer must channel his or her focus on one aspect singled out from the many possible sights or sounds present. For Proto Oceanic, this is done primarily as the object of a transitive verb. In contrast, for *smell* and *taste* the sensation is likely to be the only one of that kind available to the experiencer at that moment. As with *feel*, it is likely that we know already what we are focusing on, particularly if we are in contact with the object perceived. So in Proto Oceanic it is more usual with *smell* and *taste* for the source to be the subject of an intransitive verb, if necessary with a qualifier.

Mention was made earlier of the possibility of a universal hierarchy within which the senses are ordered, which will predict the direction of semantic change. Viberg (1984) finds some evidence for the hierarchy sight > hearing > touch > smell, taste. Comparison

of a large sample of Oceanic languages shows that most verbs of sensing have remained dedicated to a single sense. For most people, sight is the primary source of objective data about the world, and evidently was treated as such by Proto Oceanic speakers. We have no examples from a sample of many dozens of languages where a verb meaning 'see' has extended its meaning to other senses, although it can carry a cognitive meaning like 'know' or 'recognise'.

In contrast, *roŋoR 'hear' is the most semantically elastic of the sense terms. In some languages of the Solomons, Vanuatu and Polynesia, reflexes, still with the primary meaning 'hear', can be extended to 'smell', 'taste' and 'feel', although never to 'see'. In the following (admittedly very small) sample of eight languages (two from Southeast Solomonic, two from North Central Vanuatu and four from Polynesia) it can be seen that, besides *hearing*, the bundles of senses included are *hearing*, *smell* and *taste* (4), *hearing* and *touch* (2) and *hearing* and *taste* (1). There are no cases where *touch* is grouped together with *smell* and *taste* while excluding *hearing*. Thus, if the descriptions are accurate, it is noteworthy that the pattern of semantic extensions does not correspond exactly to Viberg's hierarchy in that *see* remains outside the hierarchy, while in several languages *smell* and *taste* outrank *touch*.

Table 3: Sense extensions of reflexes of POc *roŋoR 'hear' in some Oceanic languages

	see	hear	smell	taste	touch
Lau	–	y	y	–	–
Gela	–	y	–	–	y
Mota	–	y	y	y	y
Raga	–	y	–	–	y
Samoan	–	y	y	y	–
Tikopian	–	y	y	y	y
Rennellese	–	y	–	y	–
Maori	–	y	y	y	y

Reflexes of POc *ñami-, PPn *namu 'to taste s.t.', have evidently undergone a shift in meaning in a number of Polynesian languages. Some reflexes now refer to odour as well as flavour, and the shift is complete in Pukapukan, Rennellese, Samoan and Tikopia, where reflexes refer to odour alone. In view of this example it is possible that *taste* should precede *smell* in the hierarchy, although Viberg brackets the two together.

A number of commentators have proposed that, when it comes to extending sensory verbs to refer to cognitive processes, humans are visual-centric. That is, verbs of cognition, like 'know', 'think' and 'understand', tend to be recruited from verbs of seeing. However, Evans and Wilkins (2000:549) write that

> in Australian languages it is hearing, not vision, that regularly extends into the cognitive domain, going beyond the expected extension of 'hear' to 'understand' and on to 'know' ... and other cognitive verbs.

This contrasts with the Indo-European based pattern described by Sweetser (1990) in which vision is the precursor of knowing. Reflexes of POc verbs *kita and *re(k,q)i 'see' and *qilo 'be aware of, discern, see', indicate that, as in the Indo-European pattern, 'know' is more closely affiliated with 'see' than 'hear'. On the other hand, we have examples

where 'understand' is an extension of both 'see' and 'hear'. We do not have a large enough sample to draw a conclusion. It may be that context permits either.

Evans and Wilkins (2000:567) also describe the extension of meaning from 'hear' to 'obey' as common in Australian languages. We have examples of the same link from 'hear' to 'listen' to 'obey' across a number of subgroups (Southeast Solomonic, Micronesian, North and Central Vanuatu, Fijian, Polynesian). Sweetser (1990:42) writes that it is widely attested in Indo-European languages and suggests that the link may well be universal.

References

Bender, Byron, Ward H. Goodenough, Frederick Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-Min Sohn, Stephen Trussel and Judith Wang. 2003. Proto-Micronesian reconstructions. *Oceanic Linguistics* 42(1).1–110.

Blust, Robert. 1977. A rediscovered Austronesian comparative paradigm. *Oceanic Linguistics* 16.1–51.

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

—— 1989. Austronesian etymologies IV. *Oceanic Linguistics* 28(2).111–180.

—— 1995. ACD: Austronesian comparative dictionary. Electronic manuscript. Honolulu: University of Hawai'i.

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

Clark, Ross. 2009. *Leo Tuai. A comparative lexical study of North and Central Vanuatu languages.* Canberra: Pacific Linguistics.

Clark, Ross and Bruce Biggs, 2006. POLLEX: Polynesian lexicon. Electronic files. Auckland: University of Auckland.

Evans, Bethwyn. 2003. *A study of valency-changing devices in Proto Oceanic*. Canberra: Pacific Linguistics.

Evans, Nicholas and David Wilkins. 2000. In the mind's ear: the semantic extensions of perception verbs in Australian languages. *Language* 76(3).546–592.

Geraghty, Paul. 1986. The sound system of Proto Central Pacific. In Lois Carrington and S.A. Wurm, eds *Focal II: Papers from the Fourth International Conference on Austronesian Linguistics,* 289–312. Canberra: Pacific Linguistics.

Lynch, John, 2001. *The linguistic history of Southern Vanuatu*. Canberra: Pacific Linguistics.

Osmond, Meredith, Andrew Pawley and Malcolm Ross. 2003. Landscape. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic*, vol. 2: *The physical environment*, 35–86. Canberra: Pacific Linguistics.

Ross, Malcolm. 1998. Phonology and grammar. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic*, vol. 1: *Material culture,* 15–35. Canberra: Pacific Linguistics.

—— 2003. Talking about space. In Malcolm Ross, Andrew Pawley and Meredith Osmond, eds *The lexicon of Proto Oceanic*, vol. 2: *The physical environment*, 220–283. Canberra: Pacific Linguistics.

—— 2004. The grammaticization of directional verbs in Oceanic languages. In Isabelle Bril and Françoise Ozanne-Rivierre, eds *Complex predicates in Oceanic languages. Studies in the dynamics of binding and boundness,* 297–330. Berlin: Mouton de Gruyter.

Sweetser, Eve. 1990. *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Tryon, Darrell T. 1976. *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.

Viberg, Åke. 1984. The verbs of perception: a typological study. In Brian Butterworth, Bernard Comrie and Östen Dahl, eds *Explanations for language universals*, 123–162. Berlin: Mouton de Gruyter.

23 *The reconstruction of a dual pronoun to Proto Malayo-Polynesian*

LAWRENCE A. REID

1 Introduction[1]

The presence of first person dual pronouns in contrast with first person inclusive pronouns in many Western Malayo-Polynesian languages such as those in the Philippines is common knowledge. Cysouw (2003:154) labels such systems as 'Maranao-type' and renames 'dual' and 'inclusive' as 'minimal inclusive (1+2),' and 'augmented inclusive (1+2+3),' respectively, noting their widespread occurrence in other language families.[2]

The form of the dual pronoun, especially in many Philippine languages, *kita* 'NEUT 1+2',[3] *=ta* 'genitive 1+2', compared with reconstructed Proto Austronesian *ita 'NEUT 1+2+3' (Ross 2006) suggests that the dual pronoun was originally a first person inclusive pronoun, the reference of which has been restricted to a single first person and a single second person. In these languages, the form of the first person inclusive pronoun typically has an additional formative, labeled here as an EXPANDER, the shape of which appears to be cognate with either a first person singular pronoun (*=ku*), a second person singular pronoun (*=mu*), a second person plural pronoun (*=yu*), or a third person plural pronoun (*=da*). The irregular distribution of these added formatives has led to the conclusion that there is no possible form that can be reconstructed to their parent language, and that their development in today's languages is the result of convergent development, or drift. This paper reaches a different conclusion. Evidence suggests that a distinction between first person dual and first person

[1] It is a privilege to be able to contribute to this volume in honour of Malcolm Ross, a man widely recognised as both a scholar and a gentleman of the highest order. His keen attention to detail and wide familiarity with the grammatical systems of Western Austronesian languages have provided a continual stimulus to my own work, and to the students who have had the good fortune to have him as their advisor.

[2] According to Cysouw such systems are found also in the non-Pama-Nyungan languages of Australia, one language in southeastern New Guinea, also in the East Papuan languages Santa Cruz and Nanngu. They are also found frequently in Africa, mainly in Cameroon and Nigeria, both in the Niger-Congo and Chadic families, and also in some languages of the West Coast of the USA (California and Oregon).

[3] Ross' term 'neutral' is labeled 'un-casemarked' in some published works, and is commonly labeled 'nominative', 'independent', 'free' or 'long form' in others. The dual form *kita* 'you (SG) and I' found in many Philippine languages is not to be confused with the common Tagalog compound pronoun *kita* 'I (GEN) – you (NOM)' as in Tag *iniibig kita* 'I love you.' I wish to thank John Wolff and Hsiu-chuan Liao for commenting on an earlier version of this paper.

inclusive plural pronouns (labeled hence forth as 1DU and 1IPL, respectively) **is** reconstructable to the parent of all Philippine and other Austronesian languages outside of Taiwan, and that clear paths of development can be shown for each of the languages that maintain the distinction.

2 Proto Malayo-Polynesian Pronominal Reconstructions

To date there has only been one systematic attempt to provide a comprehensive account of the reconstruction and development of the pronominal system of the parent language of all Philippine and other Austronesian languages outside Taiwan (Plessis 1996). Plessis refers to this language as both Proto Philippine (PPH) and Proto Malayo-Polynesian (PMP),[4] implying that data from other Malayo-Polynesian languages do not affect the reconstructions that are based on Philippine language data alone (see Table 1a-b).

Table 1a: The Proto Philippine and Proto Malayo-Polynesian pronominal system (adapted from Plessis 1996:89)[a]

	THEME1	THEME2	RELATIONAL		
			Independent	Preposed	Postposed
1	i-aku	aku	n-akən	akən	-ku
2	i-kaw	ka[w]	n-imu	imu	-mu
3	[i-ia]	Ø, [ia]	n-iya	iya	-na
12(2)	i-kita	kita	n-atən	atən	-ta
122	i-kita[mu][yu]	kita[mu][yu]	(n-atən[mu][yu])	(atən[mu][yu])	-ta[mu][yu])
13	i-kami	kami	n-amən	amən	-mi
22	i-ka[mu][yu]	ka[mu][yu]	n-i[mu]yu	i[mu]yu	-[mu]yu
33	[i-ida]	Ø, [da]	n-ida	ida	-da

[a] The labels of these tables are originally in French.

Table 1b: The Proto Philippine and Proto Malayo-Polynesian pronominal system (adapted from Plessis 1996:89)

	REFERENTIAL		
	[+dynamic] [-proximate]	[-dynamic] [-proximate]	[+dynamic] [+proximate]
1	si-akən	di-akən	kV-akən
2	(si-imu)	di-imu	kV-imu
3	si-ia	di-ia	kV(n)-ia
12(2)	(si-atən)	di-atən	kV-atən
122	(si-atən[mu][yu])	(di-atən[mu][yu])	kV-atən[mu][yu]
13	(si-amən)	di-amən	kV-amən
22	(si-imuyu)	di-imuyu	kV-i[mu]yu
33	n-ida	di-ida	kV(n)-ida

[4] In recent publications I have referred to this parent language as Proto Extra-Formosan (PEF), and still consider it to be the most appropriate label for the immediate parent of all Philippine and other Austronesian languages outside of Taiwan (including also Yami, politically part of Taiwan, but clearly a member of the Bashiic subgroup of Philippine languages). However, in deference to the honoree of this volume, who disagrees with my use of the term, I forgo it for now.

Plessis reconstructs a single form *-ta for both 1DU and 1IPL ('postposed relational') forms, but adds [mu][yu] to the latter form, implying that it may have been possible to distinguish the pronouns by adding one or the other (or both?) of the two syllables to the base.

More recently, Malcolm Ross (2006) has provided a comprehensive account of the reconstruction and development of PAn pronouns, including a 'very tentative' reconstruction of PMP pronouns (see Table 2).[5]

Table 2: Ross' reconstruction of Proto Malayo-Polynesian pronominal forms (adapted from Ross 2006:542)

	NEUT	NOM1	NOM2	GEN1	GEN2	PSR
1SG	i-aku	aku	=(h)aku	=ku	=n(a)ku	[y]akən
2SG	ikahu	(i)kahu	=ka(hu)	=mu	=nihu	imu, ihu
3SG	siya	iya	=∅, =ya	=ya	=niya	—
1IPL	i-kita, ita	kita, i-ta	=ta	=ta	—	[y]atən
1EPL	i-kami	kami	=kami	=mi	=mami,	[y]amən
2PL	i-kamu	kamu	=kamu	—	—	—
	i-ka-ihu, kamu-ihu	ka-ihu	=ka-ihu, =kamu-ihu	=ihu, -mu-ihu	=nihu	hu, inihu, imu-ihu
3PL	—	sida	=da	=da	=nida	—

Ross' reconstructions are supplemented by proposals regarding the innovations which distinguish his PMP reconstructions from the system he reconstructs for PAn. He does not make a distinction between 1DU and 1IPL forms. Systematic reconstruction has also been done of the pronominal systems of a number of the generally recognised subgroups within the Philippines: Bashiic (Plessis 1996), Cordilleran (Reid 1979), Northern Cordilleran (Tharp 1974), Southern Cordilleran (Himes 1998), Sambalic (Plessis 1996), Danaw (Allison 1979), Manobo (Harmon 1979), and Bisayan (Zorc 1977), and extensive pronominal data is available for most Philippine languages. Philippine language data have also been used to support individual forms that have been reconstructed for Proto Austronesian (PAn) and PMP (Dempwolff 1938; Dyen 1974; Dahl 1976; Blust 1977).

The most recent study which examines Philippine pronominal systems is Liao (2008). In this carefully researched paper, Liao notes that while all accounts of PMP or PAn pronominal systems distinguish between first person inclusive (1IPL) and exclusive (1EPL) forms, none reconstructs a distinct first person dual pronoun (1DU) for either protolanguage, even though dual pronouns are widespread throughout Western Malayo-Polynesian languages.[6] Her paper provides a typology of first person dual pronouns in Philippine languages with the aim of determining whether or not a distinct dual form is reconstructable to the parent of these languages. Liao (2008:6) states that:

> [b]ased on data that are available to me, *no* distinct 1D pronoun forms are found in the following groups: (i) Bashiic, (ii) Inati, (iii) Kalamian, and (iv) the Subanun group of the Greater Central Philippine microgroup

---

5 The table is rotated from its original orientation to provide ease of comparison with other tables in this paper.

6 Apparently Liao did not have access to Plessis (1996) which suggests a possible distinction between PMP 1DU and 1IPL forms.

and concludes, based on the distribution of the different innovated formatives in the 1IPL form that appear to have developed when the earlier inclusive form was restricted to marking only dual pronouns, that they are not reconstructable to PMP and developed in the languages that have them as a result of convergent development or drift. In Reid (1979:260), I also suggested the possibility that the parent language of the Philippines may not have had a distinct dual form, although it is probable that a dual form existed in Proto Northern Luzon [=Proto Cordilleran]. I have now, however, come to the conclusion that PMP (and Proto Northern Luzon) must have had a distinction between dual and inclusive plural pronouns and that those languages that no longer show the distinction have lost it.

In order to limit the scope of this paper, I shall restrict my discussion solely to the formally simpler, typically enclitic, genitive and nominative pronouns, since if a distinction between 1DU and 1IPL exists in these sets, it will usually be found with similar marking in other sets. I shall draw on data from other sets only when they are relevant to the discussion at hand. It should be noted that Liao's paper contains an extensive list of tables containing Philippine pronominal data and the published (and unpublished) sources from which they are drawn. The less accessible Plessis (1996) also contains extensive pronominal tables from Philippine and non-Philippine Austronesian languages.

My tentative reconstruction of PMP first and second person genitive and nominative clitic pronouns is presented in Table 3. There are a number of questions that arise from Ross' PMP reconstructions, apart from whether or not the language had a first person dual pronoun, but these need to be addressed in a separate paper.

Table 3: Proto Malayo-Polynesian first and second person clitic pronouns[a]

	GENITIVE	NOMINATIVE
1SG	=ku	=aku
1DU	=ta	=ta
2SG	=mu	=kaʔu
1IPL	=ta[mu]	=ta[kamu]
1EPL	=mi	=kami
2PL	=mu[yu]	=kamu

[a] Square brackets indicate dialectally distinguished pronominal formatives.

2.1 First person dual and inclusive pronouns in Philippine languages

The development of a distinction between 1DU and 1IPL forms has been addressed by Blust (2009:309). He suggests that (cited by Liao):

> … [t]he most likely explanation of these special dual forms arises from the pragmatics of the speech act: most conversations take place between a speaker and a single hearer. As a result, the use of an inclusive pronoun would normally involve only the conversational dyad of speaker and hearer, whereas this would not necessarily hold for the corresponding exclusive form, since speakers commonly refer to themselves and *others* rather than a single *other*. Frequency of usage alone would lead reflexes of *kita to become de facto duals, creating a need for new plural inclusive forms, which were then cobbled together from the existing reflex of *kita plus parts of other pronouns (*-ihu*, *n-ihu*, *-m(u)* '2SG', *-da* '3PL', etc.).

Liao (2008) provides full details of the ways each Philippine language has supposedly 'cobbled together' parts of existing pronouns in order to form the new 1IPL forms. These are summarised as follows:

1. In Central Cordilleran languages the form appears to be first singular (1SG), *=ta '1DU' + *=ku '1SG' > =*taku* '1IPL'.
2. In Central Luzon languages, Northern and Southern Alta, some of the Northern Cordilleran languages, as well as Arta, and Umiray Dumaget, the form appears to be second singular (2SG), *=ta '1DU' + *=mu '2SG' > =*tamu*, =*tam* '1IPL'.
3. In the Danaw languages, Itneg, Southern Cordilleran languages, Ilokano, Tausug, Tagalog, some Manobo languages, Aborlan Tagbanwa, Palawan Batak, Tboli, etc., the form appears to be second plural (2PL), *=ta '1DU' + *=yu '2PL' > =*tayu* '1IPL'.[7]
4. In the Northern Cordilleran languages, Isnag, Malaweg, and Itawis, the form appears to be third plural (3PL), *=ta '1DU' + *=(i)da '3PL' > =*tada, =tida, =téra* '1IPL'.
5. Other means have also been used to form the distinction, such as vowel alternation in Blaan: *ta* '1DU' versus *to* '1IPL'.

While Blust is no doubt correct in his explanation of why a distinction between these two forms developed, I believe he is incorrect in assuming that languages simply chose to randomly cobble together forms from parts of other pronouns to create new forms. The fundamental question arises, why would a language choose a singular form such as 'I' or 'you (SG)' to extend the function of a dual pronoun to an inclusive one? There seems to be no available pragmatic or semantic explanation. The semantics of an inclusive first person pronoun ('we all') is not achieved by adding 'I' or 'you (SG)' to 'we (two)'.

Probably, the most widespread of the expanders used for forming the 1IPL (genitive) pronoun is what appears to be a reflex of PMP *=mu 'GEN.2SG'. Although a semantically unexpected ending, it appears not only in many of the Negrito languages noted for their conservative morphology (such as Arta, a first order branch of the Northern Luzon family and the Altan languages, two very different languages forming a sister branch with the Central and Southern Cordilleran languages in Northern Luzon) (Reid 1989, 1991), and all the Negrito languages of North-East Luzon, including Umiray Dumaget (possibly a Central Philippine language (Himes 2002)), as well as in some of their sister languages in the Cagayan Valley branch of Northern Cordilleran. It appears also in the Ayta Negrito languages and all of the other members of the Central Luzon family, including Sinauna and Kapampangan. Although Ivatan in the far north of the Philippines does not distinguish first person dual from inclusive (as noted by Liao), the distinction does exist in Yami with the inclusive pronoun being commonly formed in the same way (see §2.4 below). It is also the expander in all the languages of Mindoro, including the northern group, Tadyawan, Alangan and Iraya, as well as the genetically dissimilar southern group, including Hanunoo

[7] Various other forms appear, as noted in the Blust quote, differing according to the actual development of second person plural forms in the languages, thus for example, the Aborlan Tagbanwa 1IPL form is =*tami*, where =*mi* is not a 'cobbled' 1EPL form, but is the genitive 2PL pronoun that developed from *=muyu (see §2.2 below), and Brooke's Point Palawano is =*tiu*, apparently by analogy with Brooke's Point Palawano =*miu* 'GEN.2PL'.

and Buhid. In the far south of the Philippines, Tiruray, a member of the Southern Mindanao family distinguishes two of the first person genitive pronouns ('Set C') as follows: *to* '1DU' versus *tom* '1IPL' (Schlegel 1971), where /o/ is one of the Tiruray reflexes of PAn *a (Blust 1992).

The other semantically unexpected expander, the apparent reflex of first person singular *=ku is relatively restricted, occurring in most of the fairly closely related set of Central Cordilleran languages in northern Luzon.

The other widespread expander used for forming the 1IPL pronoun is a not unexpected second person plural pronoun, forming the semantics of 'I, you (SG), and the rest of you'. It is often a reflex of a genitive *=yu, or *=niyu. These appear in some of the northern languages of the Philippines, such as Ilokano and the various Itneg languages,[8] all of the Southern Cordilleran languages, a few of the Central Philippine languages, such as Tagalog and Tausug (as *=niyu*),[9] in the Danaw languages (as *=nu*), in some of the Manobo languages, and in some of the languages of Palawan, such as Palawano, Aborlan Tagbanwa and Batak.

There are a few languages, noted above, in which the expander is a third person plural form. This is not unexpected in that it produces the semantics of 'I, you (SG), and the others.'

The major problem then, lies in finding some explanation for the most widespread of the expanders. Under the explanation proposed by Blust (and endorsed by Liao) that the distinction between first person dual and inclusive pronouns is not reconstructable to the parent language of the Philippines, one must assume that a semantically inappropriate second person singular form was independently chosen from among the various available pronouns as the most appropriate ending for the inclusive form, not once, but numerous times, as it occurs in a number of primary subgroups or individual languages from one end of the Philippines to the other. This, to me, seems highly unlikely.

2.2 On the development of *=tamu* forms in Malayo-Polynesian languages

In an oft-cited paper, Blust (1977) demonstrates that PMP *=mu 'GEN.2SG' reflects the PAn clitic *=mu '2PL', replacing the earlier PAn *=Su 'GEN.2SG', a change that constitutes one of the defining innovations of his PMP and which he labeled as 'the second politeness shift' (see also Ross 2006:541–542). If we assume that *=ta became restricted to 1DU with the concurrent development of a new 1IPL form in PMP *prior* to the shift of the second person genitive pronoun from plural to singular, we have a reasonable explanation for *=mu 'GEN.2PL' being chosen as the combining form for the new pronoun *=ta[mu] 'GEN.1IPL', and we have an explanation for its widespread occurrence in Philippine languages. These are inherited forms, not innovations.

With the shift of meaning of *=mu from second person plural to second person singular and the breakup of the parent language and dispersal of its daughters, widespread reformation of the 1IPL pronoun occurred. The nature of these changes suggests that the combined form was not fully lexicalised in PMP, that is, for some native speakers the expanded form was interpreted as a single morpheme, while for others the expander was

---

8 Itneg is a closely related group of Central Cordilleran languages strongly influenced by Ilokano.

9 The modern Tagalog 1IPL form *=táyo* is a nominative clitic and is distinct from an old nominative dual *=kata*, still used in some rural areas. The genitive 1IPL form *nátin* is used today in contrast to dual *=nita* (Schachter and Otanes 1972:88).

interpreted as a second morpheme. Variation also probably occurred,[10] with some dialect areas using the combined form while others did not, a situation not unlike the *you*/*you [all]* variation of the second person pronoun in English. This may account for the few subgroups which still today do not have any contrast between dual and inclusive first person forms. There is clear evidence, however, that some languages have lost the contrast in relatively recent times. Although dual forms are known for some dialects of Tagalog, for example, the historically newer first person inclusive forms are commonly used today for both 1DU and 1IPL. Similarly, although there is no contrast between the two forms in Ivatan, its sister Bashiic language, Yami in Taiwan, maintains the contrast (see §2.4).

Languages that still retain the full form *=tamu* (including all of the Central Luzon languages) appear to have fully lexicalised the pronoun, in that they have not substituted the semantically inappropriate second person singular ending with a more appropriate plural ending. Others maintain it, but have reduced it to *=tam.* In the Negrito languages of Northeastern Luzon, the Altan languages, and several of the Cagayan Valley languages this could have developed by analogy with the genitive second person singular variant *=m* that in all these languages occurs after vowel-final words. But the shortened form *=tam* also occurs in languages such as Arta, and Casiguran Dumagat that do not have a shortened second person singular variant *=m* following vowel-final words, implying that, at least in these languages, the final consonant of the form is not identified as a second person singular ending. Tiruray in the south of Mindanao also has a reduced form *=tom*, but only has *=mu* as the second person singular form. I consider it most likely that in all the languages that have the shortened form *=tam*, it is the result of vowel erosion, rather than by analogy with the shortened form *-m* of the genitive second person pronoun.

Languages that inherited *=ta[mu] as a sequence of meaningful parts, rather than as a single lexicalised pronoun, replaced the inappropriate ending following its shift from second person plural to second person singular. The most common substitution was with the form *=yu*, a second person plural pronoun that possibly developed after *=mu changed its meaning. The source of this form is of some interest and requires us to evaluate the reconstructions that have been made, specifically those proposed by Ross (2006). He lists the set of changes related to Blust's 'second politeness shift' as follows:

a. The PAn plain neutral *i-Su 2SG is lost, and PMP *ikahu, reflex of PAN polite *i-ka-Su, becomes the default neutral 2SG pronoun.

b. PMP *=mu GEN.2SG reflects the PAn clitic *=mu 2PL, and the PAn clitic *=Su 2SG is lost (although the long clitic *=nihu, reflex of PAn *(=)ni-Su continues).

c. PMP has new additional forms, neutral *[i]ka-ihu and *kamu-ihu and genitive *=ihu, *=nihu, *=mu-ihu which incorporates *-ihu, apparently reflecting PAn neutral *i-Su 2SG. (Ross 2006:542)

While this set of changes provides a 'top-down' account for the proposed PMP reconstructed forms suggesting the paths of development from earlier PAn forms, problems arise when attempting to account for their reflexes in Philippine languages. The first problem has to do with the forms; the second problem has to do with their function.

It is a well-known fact of PMP phonology[11] that PAn *S > PMP *h and that *h was lost in all languages of the Northern Luzon subgroup,[12] possibly accounting for the

[10] Hence its representation as PMP *=ta[mu], rather than *=tamu.

[11] 'It is clear that by PMP times the sibilant reflexes of *S were completely lost outside of Formosa' (Zorc 1982:121).

development in these languages of a new second person plural genitive *=yu* (< *=ihu). But PMP *h was retained in other Philippine subgroups, so we would expect to find in these other languages at least some second person plural genitive forms with the shape *=ihu*, *=nihu* or *=mu-ihu*. We do not. What we find instead are forms without *h*, such as *=yu, =niyu, =muyu* and the like. Such forms are found throughout the languages of Central Luzon, Central Philippines, Manobo, Danao, and other Philippine subgroups. *=muyu* is also found in languages in Borneo, while the analogically changed *=miyu,*[13] *=myu* or *=miu* is found in Itbayaten, Brooke's Point Palawano[14] and Northern Alta in the Philippines, as well as in Chamorro and Palauan, some of the languages of Sulawesi, Eastern Indonesian and Oceania (Harvey 1982; Plessis 1996:78). The shortened form *=mi* 'GEN.2PL' is found in Aborlan Tagbanwa,[15] and also some Sulawesi languages (e.g., Uma and Da'a) (Himmelmann 1996). Either we need to claim that PMP *h was independently lost in just these forms in the parent of each of these subgroups or that the reconstructions tentatively claimed in Ross (2006:542) and originally proposed by Blust (1977) need to be revised.

The second problem with the reconstructions that Ross proposes is found in his suggestion that the new PMP genitive forms (*=ihu, *=nihu, and *=mu-ihu) (following Blust 1977) apparently reflect PAn *iSu[qu] 'NEUT.2SG'. This implies that second person singular and plural forms each reversed their number feature in PMP: PAn *Su 'GEN.2SG' was replaced in PMP by *=mu (formerly 'GEN.2PL'), while PAn *=mu 'GEN.2PL' was replaced in PMP by *-ihu (supposedly reflecting PAn *-iSu[qu] 'NEUT.2SG'). The first replacement is accounted for by Blust's 'second politeness shift',[16] but no account is available for the second replacement. Moreover, we are faced with questions of the same type that initiated this enquiry, if *-iSu was a singular pronoun, why would it have been added to *=mu '2SG' to form a plural form, *=mu-ihu '2PL'?

Where Ross reconstructs PMP *=mu-ihu 'GEN.2PL', Plessis (1996:77), considering the distribution of its apparent reflexes (noted above), reconstructs PMP *=muyu 'GEN.2PL' while accepting as plausible the claim that the second syllable had its source in PAn *-iSu. 'Selon Blust (1977), *-muyu serait formé en réalité de 2 morphèmes *-mu et *-iSu. Les correspondances phonologique entre PAn et PMP rendent cette hypothèse plausible.' Plessis nevertheless notes that Harvey (1982:79) rejects *-iSu as the source of PMP *=yu, on the basis of the appearance of a possible cognate in Saisiyat *moyo* 'GEN.2PL'.[17]

Whatever the origin and function of the final syllable of Saisiyat *moyo*, it is apparent that a form *=yu occurred in PMP to sustain a distinction between singular and plural second person forms, PMP *=mu '2SG'/*=mu[yu] '2PL' during the period following loss of the earlier PAn *Su 'GEN.2SG'. Plessis suggests that the genitive second person plural *=yu* found in Central Cordilleran languages developed from either loss of the first syllable

---

[12] For example: PMP *húRas > Tagalog *húgas*, but Batad Ifugao *ūlah* 'wash'; PMP *buhek > Tagalog *buhúk*, but Bontok *fuʔúk* 'hair', PMP *tebuh 'sugarcane' > Tagalog *tubúh-an* 'sugarcane field, plantation', but Bontok *túfu* 'leaf, plant', etc.

[13] The initial vowel appears to have been changed by analogy with the genitive first person exclusive pronoun *=mi*.

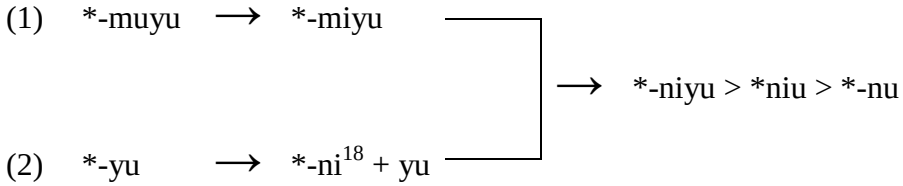
[14] Information courtesy of Bill Davis.

[15] Data provided by Neva Bisquer, a native speaker of the language.

[16] Wolff (2007:15–16) rejects Blust's speculative account of the change of meaning of PAn *mu 'GEN.2PL' to PMP *mu 'GEN.2SG', claiming instead that the forms are the result of regular phonological developments which have resulted in homophony. The issue is not directly relevant to the issue at hand and will not be pursued further here.

[17] PAn *S is typically reflected in Saisiyat as /ʃ/, not /y/.

(*=muyu > *=[mu]yu > *=yu), or by phonological reduction (*=muyu > *=myu > *=yu). The possible developments in PMP languages are charted by Plessis (1996:79), ending with *=nu*, the genitive 2PL form found in a number of the Cagayan Valley languages of Northern Luzon, as follows:



The shift of the genitive second person plural form from *=mu[yu] to *=yu seems to have been paralleled by the shift of the genitive first person inclusive plural from *ta[mu] to *ta[yu],[19] both apparently motivated by the inappropriateness of an explicitly second person singular form constituting part of an explicitly plural form, rather than simply by phonological reduction as proposed by Plessis.

2.3 On the development of *=taku* forms in Central Cordilleran languages

While the development of *=tamu and *=tayu forms follows naturally from the reconstruction of PMP *=ta[mu] '1IPL', and *=mu[yu] '2PL', an explanation of reflexes of what has been reconstructed as Proto Central Cordilleran (PCCo) *=taku 'NOM/GEN.1IPL' (Reid 1974, 1979) is needed. The claim being made here is that *=taku* forms are **not** the result of the cobbling together of a dual form *=ta* and a first person singular form *=ku*, as has been repeatedly claimed in the literature, but rather the expander *=ku* is a reduced form of the nominative second person plural clitic *=kayu*. Evidence for this claim comes from several languages.

In Philippine languages nominative second person **singular** clitic forms are typically reflexes of a PMP *=kaʔu > *=kaw, as in TAG *ʔikaw* '2SG'. Ross (2006:542) reconstructs it as PMP *=ka(hu),[20] however no forms are found outside of Taiwan with a reflex of a medial *h.[21] That the form was disyllabic, rather than monosyllabic *ka[w] as reconstructed by

---

[18] The form *ni is assumed by Plessis to be what is commonly referred to as the genitive personal noun marker.

[19] As in all the Southern Cordilleran languages, such as Ilongot, Pangasinan, Inibaloi, etc., as well as in Ilokano and Itneg, e.g., Ilk *áso=tayó* 'our (1IPL) dog'.

[20] Ross (2006:524) indicates that forms in parentheses either did, or did not occur.

[21] Bonggi, a language of Sabah, Malaysia, which according to the Ethnologue (Gordon 2005) is one of the Palawano languages, closely related to Molbog of Balabac Island in the south of Palawan in the Philippines, has a medial *h* in several pronouns, including: *uhu* 'NOM.2PL' and *dihu* 'DAT/ACC.2PL' (Boutin 1988:5, 2002:215). It is clear however that Bonggi /h/ has developed from PMP *k, and not from PAN *S. The published data suggest that voiced and voiceless bilabial and velar stops have fricative variants as evidenced from the pronominal alternations recorded in Boutin (1988:5): *ku/hu* 'GEN.1SG', *kita/hita* 'NOM.1DL' *kiti/hiti* 'NOM.1IPL'. Kroeger (1992:295) notes that 'for many speakers, /k/ does not spirantise to [h] in stressed syllables'. At the onset of unstressed syllables, however, it seems the change is complete, as in the second person pronouns cited above. The evidence suggests that the sound change is still in progress in initial (stressed) position. In addition to the consonants, regular sound changes relate the final syllables in Molbog *kitey* [kitəy] and Bonggi *kiti/hiti* 'NOM.1IPL', and Molbog *ekew* [əkəw] and Bonggi *uhu* 'NOM.2PL', while the initial *u* in the 2PL form is a result of one of the Bonggi vowel harmony rules (Kroeger 1992) by which high vowels can spread from left to right, or right to left, as in /dabu/ + -an > [du'βuuan] 'fall'.

Plessis (1996), is supported by the presence of disyllabic forms in Chamorro as well as some of the lesser known languages of Mindoro in the Philippines, such as Alangan *kaʔu*, Hanunoo and Iraya *kawu* 'NOM.2SG', as well as in a number of languages south of the Philippines, with forms such as Minangkabau and Banjarese Malay *kau* '2SG' (Adelaar 1992:113). The reconstruction of a medial glottal stop in PMP *=kaʔu, is controversial. That the form was disyllabic is clear, but whether the onset of the second syllable was a glottal stop (as in other reconstructed PMP and Proto Philippine forms, such as *laʔun 'long; slow', *haːʔun 'remove from fire', *baʔug 'rotten (eggs), sterile', etc. (Zorc 1982:128–129, 1996)), or some other transitional consonant, such as *w, following the loss of an earlier consonant, or whether there was no consonantal onset to the second syllable is unclear. I assume however that the sequence *aCu > *a[ʔ]u > *a[w]u > *a[Ø]u > *aw* is more natural than *aCu > *a[w]u > *a[ʔ]u > *a[Ø]u > *aw*,[22] or any other permutation of these changes.

The equivalent **plural** form was PMP *=kamu 'NOM.2PL', a form that, like PMP *=tamu 'GEN.1IPL', was not fully lexicalised in that following the shift of *=mu to second person singular, it was replaced with the 'appropriate' plural form *=yu, resulting in *=kayu* 'NOM.2PL' in many languages.

In many Philippine languages, *=ta* '1DU' is both a nominative and a genitive pronoun, similarly *=taku* and *=tayu* '1IPL' are both nominative and genitive pronouns. Evidence suggests, however, that nominative and genitive forms were originally distinct, that is *=tayu* '1IPL' was formed with a combination of *=ta* and a **genitive** second person plural *=yu*, while *=taku* was formed with a combination of *=ta* and a **nominative** second person plural *=kayu*, each of which subsequently extended their function to cover both cases. The presence of a reflex of *=taku 'NOM.1IPL' has generally been considered to be one of the distinguishing features of the Central Cordilleran subgroup of Northern Luzon languages, but there is one Central Cordilleran language, Kalinga, as spoken in Manabo, Abra that maintains a reflex of what must be an earlier form *=ta[kayu] 'NOM.1IPL' as shown in Table 4.[23]

Table 4: Manabo Kalinga personal pronouns
(adapted from McFarland 1977:13)

	GEN	NOM
1SG	=ku/=k	=ʔak
1DU	=ta	=ta
2SG	=nu/-m	=ka
3SG	=na	siya
1EPL	=mi	=kam
1IPL	=takay	=takay
2PL	=yu	=kay

[22] The symbols in square brackets indicate transitional consonants which may or may not have had phonemic status in the language.

[23] Manabo Kalinga has lost final high vowels in what were originally di- or tri-syllabic pronouns. It should be noted also that *=nu* 'GEN.2SG' is probably analogically changed from earlier *=mu 'GEN.2SG', to match the alveolar initial consonant of the third singular pronoun (*=na* 'GEN.3SG'), and thus has a different source than *=nu* 'GEN.2PL' found in some Cagayan Valley languages that appears to have developed from earlier *=niyu. The reduced form *-m* of *=mu 'GEN.2SG' is retained as a second person singular agreement feature on vowel-final words in Manabo (and other Itneg-Kalinga languages) (Reid 2001).

An explanation is needed for the high vowel of the final syllable of *=taku* in other Central Cordilleran languages, if in fact it developed from *ta[kayu]. One possible explanation is that the disyllabic form was reduced by analogy with the first person singular genitive pronoun *=ku, but one might expect that if this was the case, then the pronoun would have developed as *=tak, since in all of the Central Cordilleran languages in which *=taku* occurs, *=ku* 'GEN.1SG' is reduced to *=k* on vowel final forms. A different type of phonological explanation is probably more satisfactory. Stress in all languages which have a full reflex of *=ka'yu 'NOM.2PL' is invariably on the ultimate syllable. The low vowel in the initial syllable of the form is unstressed and is susceptible to assimilation to the stressed high back vowel of the final syllable. Although there are no reflexes in any of the Central Cordilleran languages showing raising of the unstressed vowel in the 2PL form, in Batad Ifugao the vowel is lost when it is part of the medial syllable of the formerly trisyllabic neutral or nominative second person plural pronoun *daka'yu, reflected as *daʔyu* 'NEUT.2PL'.[24] Nevertheless, Ifugao does show the same assimilation in the initial weak vowel of its reflex of *=ta'ku 'NOM/GEN.1IPL', IFGBT *=tu'ʔu* 'NOM/GEN.1IPL' (see Table 5).

Table 5: Some Batad Ifugao personal pronouns
(adapted from Newell 1993:211)

	Nominative	Genitive
1SG	=aʔ	=ʔu, =ʔ
1DU	=ta, dita	=ta
2SG	=ʔa, daʔa	=mu, =m
1EPL	=ʔami, daʔmi	=mi
1IPL	=tuʔu, dituʔu	=tuʔu
2PL	=ʔayu, daʔyu	=yu

However, there are at least two languages outside of the Cordilleran area that have replaced the inherited expander on the second person plural pronoun *=kamu with the semantically appropriate second person genitive ending *=yu, giving *=ka'yu, and have then assimilated the vowel in the first syllable to that in the final. The Northern Mangyan language, Iraya, maintains the full form *=kuyu* 'NOM.2PL', while the Southern Mindanao language, Tboli, has reduced the form from earlier *=kuyu to *kuy* '2PL', and this is the form that has been added as an extender to the dual pronoun to produce the current first person dual form *tekuy* (see Table 6). The claim then is that the Central Cordilleran languages underwent the same processes of assimilation and elision that eventually resulted in the form *=taku* '1IPL'.

Table 6: Some Tboli first and second person pronouns
(adapted from Porter 1977:35–36; Forsberg 1992:22)

	NON-FOCUSED PRONOUNS	
	Set III-*u* class	Set IV-*dou* class
1DU	te	kut
2SG	-em/-hem/-m	kóm
1IPL	tekuy	tekuy
2PL	ye	kuy

[24] PAN *k regularly became /ʔ/ in Batad Ifugao.

2.4 On the development of dual first person forms in Yami

Yami, spoken on Botel Tobago Island in Taiwan, has long been recognised as a sister of Itbayaten, Ivatan and the other Bashiic languages, possibly as a result of back-migration from the Philippines within the last millennium or so. Its position outside of direct influence from Philippine languages therefore makes the fact that it has a distinction between 1DU and 1IPL pronouns of considerable relevance to the issue as to whether or not PMP had a dual pronoun. There has been no mention of a dual pronoun in any of the available materials on the language, such as Ho (1990), although Asai (1936:42) lists two first person inclusive forms, *ta* and *takamu*, and Rau and Dong (2006:123) list three ('bound') first person inclusive forms *=ta*, *=tamo* and *=takamo*, but without any discussion of the different functions of the forms. An examination of the extensive text materials for Yami that are available in Rau and Dong (2006) (henceforth RD) as well as the Yami translation of the New Testament (NT) provides clear evidence that while *=ta* is used for both dual and inclusive pronouns, extended forms are restricted to first person pronominal reference with more than two people.

There are more than 75 non-singular, first person references in the RD texts, at least 53 of which can be unambiguously identified from context as having dual reference. None of the dual pronouns uses one of the 'new' extended forms.[25] Of these dual references, *=ta* is used to encode genitive possessors (9 times) as in (1), genitive actors of a transitive construction (17 times) as in (2), and nominative actors of an intransitive construction (16 times) as in (3). The other pronouns with dual reference are encoded with independent casemarked long forms, either *yaten* for nominative patients of a transitive construction (6 times) as in (4), or as fronted sentence topics as in (5), *niaten* for the oblique patient of an extended intransitive construction (once) as in (6), and *jiaten* for the locative patient of a subjunctive transitive verb (twice) as in (7). For these independent pronominal forms, no contrast between dual and inclusive pronouns exists. In the examples which follow, the relevant pronouns are bolded.

(1) *among* ***ta*** 'our (DU) fish' (RD18.6)

(2) *mi* ***ta*** *moaen o ovi* 'let's (DU) go plant the yams' (RD14.2)

(3) *mi* ***ta*** *manazataza* 'let's (DU) go fishing' (RD18.2)

(4) *ta na* ***yaten*** *zakaten* 'because he has come to kill us (DU)' (RD1.21)

(5) ***yaten*** *rana ya am ...* 'as for us (DU) already …' (RD14.23)

(6) *o ito so manci* ***niaten*** *a rarakeh a* 'the old man is calling to us (DU)' (RD11.6)

(7) *ji abo kaji na zakatan* ***jiaten*** 'he will surely kill us (DU)' (RD16.81)

In contexts in which the first person pronoun can be unambiguously identified as referring to more than two people, the extended form *=tamo* is most commonly used in both RD and NT. It encodes the nominative 1IPL actor of an intransitive construction (7 times) as in (8)–(9) and the genitive 1IPL actor of a transitive construction (3 times) as in (10).

---

[25] There is one identified instance in the NT where St. John is writing to one person but uses *tamo* for dual reference, e.g., *Inaoy si makeykai ko rana makacita jimo a kawalawalam* ***tamo*** *rana makacita so kadwan jyaten.* 'I wish that I will soon see you (SG) and we will be together to see each other.' 3 John v.14. This has been confirmed by a native speaker as a mistranslation. The underlined form should be *kawalawalam* ***ta*** *na*, confirming the analysis presented in this section.

(8) *mangay* ***tamo*** *do Jihaod ori* 'let's (all) go to Jihaod' (RD11.10)

(9) *Meyoli* ***tamo*** *pa do Yotay a pongso*. 'Let us (all) go back to Judea.'
(John *Yohani* 11:7)

(10) *kalaen* ***tamo*** *pala sira* 'let's (all) look for them' (RD15.28)

The longer extended form, *=takamo* appears in RD only as the genitive 1IPL actor of a transitive construction (5 times) as in (11), but it is also found in NT as the nominative 1IPL actor of an intransitive construction as in (12).

(11) *kabkaben* ***takamo*** *o vahay da* 'we (all) will tear open their roof' (RD15.25)

(12) *O ney manma na nimamizing nyo a nanao am masicyakarilaw* ***takamo*** *an.*
'The very first teaching you heard was that we (all) love-one-another.'
(1 John *1 Yohani* 3:11).

While genitive actors of transitive sentences distinguish 1DU from 1IPL forms, in that the extended forms are only used to encode the latter, genitive possessors in noun phrases do not distinguish them; *=ta* is used to encode both 1DU, as in (1) above, and 1IPL. Extended forms do not encode genitive possessors; this is true for both RD and NT texts. Thus in (13), the first pronoun is the nominative actor of an extended intransitive construction, while the second pronoun is the possessor of the noun expressing the oblique patient; both are first person inclusive plurals and have the same reference, but the forms are different. Similar examples are found in (14)–(16).

(13) *mi* ***tamo*** *milolo so kayo-kayo* ***ta*** 'Let's (all) go drag our wood.' (RD6.1)

One other factor that is apparent from the NT data, is that all three forms can alternate to encode the genitive 1IPL actor of a transitive sentence, although in RD, *=ta* does not appear with this function. Examples (14)–(16) are almost identical in meaning but show different pronominal forms.

(14) *Isaray* ***ta*** *ji Ama* ***ta*** *do to a Ama na ni Yeso Kizisto a Panirsirngen* ***ta***.
'We (all) thank our Father, the Father of Jesus Christ our Savior.'
(Ephesians *Ivoso* 1:3)

(15) *Isaray* ***tamo*** *ji Ama na ni Yeso Kizisto a Panirsirngen* ***ta***.
'We (all) thank the Father of Jesus Christ our Savior.'
(2 Corinthians *2 Kedinto* 1:3)

(16) *Isaray* ***takamo*** *si Ama* ***ta*** *do to a Ama na ni Yeso Kisizto a Panirsirngen* ***ta*** *…*
'We (all) thank our Father above, the Father of Jesus Christ our Lord …'
(1 Peter *1 Pite* 1:3)

Just as it is possible to infer from some languages in the Philippines that dual forms were expanded by both genitive and nominative second person pronouns, *=yu* and *=kayu* respectively, to form plural inclusive forms, the same may have been true also for PMP, given that reflexes of both *=mu and *=kamu occur as dual expanders. There are no languages today however in which the expanded forms are distributed accordingly.

Today, theYami genitive second person plural form corresponds to that found in Ivatan, *=nyo*, from earlier *=niyu. There is at least one example found in NT which shows a reformation of the 1IPL pronoun by replacing the inherited expanders that appear to be semantically inappropriate, with the current 2PL form to create a new 1IPL pronoun, *=tanyo*, as in (17).

(17) *Cyaha, ta **tanyo** apiya.* 'Nevermind, because we (all) are fine.'
(1 Thessalonians *1 Tesalonicya* 5:3)

3 Conclusion

The recognition that PMP must have had a contrast between first person dual and inclusive plural pronouns began with the question of why a language would have chosen to add a singular first or second person pronoun as an expander to a dual form in order to reconstitute the semantics of an inclusive form. It seemed improbable that most Philippine subgroups and some non-Philippine groups speaking Malayo-Polynesian languages would have independently chosen a second person **singular** pronoun (*=mu) as the expander to create a distinction between 1DU and 1IPL pronouns, as had been claimed. Recognizing that this form was a second person **plural** pronoun in Formosan languages, accounting for its reconstruction as such to PAn, provides an explanation for its widespread occurrence as an expander forming a contrast between *=ta '1DU' and *=tamu '1IPL'. This explanation is only possible if the contrast actually existed in PMP, prior to the shift of *=mu from second person plural to second person singular (Blust's 'second politeness shift') and accounts for its wide distribution among PMP languages.

I claim that following the shift from second person plural to singular, the inappropriate semantics of *=ta+mu as a first person inclusive plural form resulted in the replacement of the expander by the new second person plural form *=yu in those daughter languages that had not fully lexicalised the combination and were thus associating the expander with the second person singular pronoun. This shift accounts for the widespread occurrence of *=tayu '1IPL'.

The claim that the first person singular pronoun (*=ku) was the source of the expander of *=ta* '1DU' to *=taku* '1IPL' in the Central Cordilleran languages was problematic not only because of the irregular semantics, but also because in all of these languages there is a vowel-less variant of *=ku* when cliticised to any vowel-final form, so that one would expect the inclusive plural pronoun to have been *=*tak,* if the first person singular pronoun had in fact been the source of the expander. The occurrence of reflexes of a nominative second person plural form *=ka'yu, in which the unstressed first vowel had assimilated to the stressed final high vowel as *=ku'yu,* along with evidence that *=ka'yu 'NOM.2PL' was the expander in one of the Central Cordilleran languages provided an explanation that this was the source of the *=ku* expander in the other languages of that subgroup.

References

Adelaar, K. Alexander. 1992. *Proto Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

Allison, E. Joe. 1979. Proto-Danaw: a comparative study of Maranaw, Magindanaw and Iranun. In *Papers in Philippine Linguistics No. 10,* 53–112. Canberra: Pacific Linguistics.

Asai, Erin. 1936. A study of the Yami language: an Indonesian language spoken on Botel Tobago Island. PhD thesis. Leiden: Universiteitsboekhandel en Antiquariaat J. Ginsberg.

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.

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

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

Boutin, Michael E. 1988. Banggi clause structures. In Charles Peck, ed. *Borneo language studies 1: Sabah syntax paper*, 1–52. Language Data, Asian-Pacific Series, 14. Dallas: Summer Institute of Linguistics.

—— 2002. Nominative and genitive case alternations in Bonggi. In Fay Wouk and Malcolm Ross, eds *The history and typology of Western Austronesian voice systems*, 209–239. Canberra: Pacific Linguistics.

Cysouw, Michael. 2003. The paradigmatic structure of person marking. PhD thesis. Nijmegen: The Catholic University of Nijmegen.

Dahl, Otto Christian. 1976. *Proto-Austronesian*. Lund: Studentlitteratur.

Dempwolff, Otto. 1938. *Vergleichende lautlehre des Austronesisches wortschatzes*, vol. 3. Berlin: Dietrich Reimer.

Dyen, Isidore. 1974. The Proto-Austronesian enclitic genitive pronouns. *Oceanic Linguistics* 13(1–2).17–31.

Forsberg, Vivian M. 1992. A pedagogical grammar of Tboli. *Studies in Philippine Linguistics* 9(1).1–110.

Gordon, Raymond G., Jr., ed. 2005. *Ethnologue: languages of the world*. 15th edition. Dallas, Texas: SIL International.

Harmon, Carol. 1979. Proto-Manobo pronouns and case marking particles. In *Papers in Philippine linguistics No. 10,* 113–133. Canberra: Pacific Linguistics.

Harvey, Mark. 1982. Subgroups in Austronesian. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics,* vol. 2, *Tracking the travellers*, 47–99. Canberra: Pacific Linguistics.

Himes, Ronald S. 1998. The Southern Cordilleran group of Philippine languages. *Oceanic Linguistics* 37(1).120–177.

—— 2002. The relationship of Umiray Dumaget to other Philippine languages. *Oceanic Linguistics* 41(2).275–294.

Himmelmann, Nikolaus P. 1996. Person marking and grammatical relations in Sulawesi. In H. Steinhauer, ed. *Papers in Austronesian Linguistics No. 3,* 115–136. Canberra: Pacific Linguistics.

Ho, Arlene Y.L. 1990. Yami structure: a descriptive study of the Yami language. MA thesis. Hsinchu: National Tsing Hua University.

Kroeger, Paul R. 1992. Vowel harmony systems in three Sabahan languages. In Peter W. Martin, ed. *Shifting patterns of language use in Borneo: papers from the second bi-ennial International Conference, Kota Kinabalu, Sabah, Malaysia July, 1992*, 279–296. Borneo Research Council Proceedings Series, 3. Kota Kinabalu: Borneo Research Council.

Liao, Hsiu-Chuan. 2008. A typology of first person dual pronouns and their reconstructibility in Philippine languages. *Oceanic Linguistics* 47(1).1–29.

McFarland, Curtis D. 1977. *Northern Philippine linguistic geography*. Study of Languages and Cultures of Asia and Africa Monograph Series. Tokyo: Tokyo University of Foreign Studies.

Newell, Leonard E. 1993. *Batad Ifugao dictionary with ethnographic notes*. Linguistic Society of the Philippines Special Monograph No.33. Manila: Linguistic Society of the Philippines.

Plessis, Frédéric. 1996. *Morphologie de Proto-philippine II: Reconstruction du système pronominal Proto-philippin (PPH)*, Mémoire de DEA. Langues, littératures et civilisations: Etudes austronésiennes. Paris: Institut National des Langues et Civilisations Orientales (INALCO).

Porter, Doris. 1977. *A Tboli grammar*. Philippine Journal of Linguistics Special Monograph No.7. Manila: Linguistic Society of the Philippines.

Rau, D. Victoria and Maa-Neu Dong. 2006. *Yami texts with reference grammar and dictionary*. Language and Linguistics Monograph Series A-10. Taipei: Institute of Linguistics, Academia Sinica.

Reid, Lawrence A. 1974. The Central Cordilleran subgroup of Philippine languages. *Oceanic Linguistics* 13.511–560.

—— 1979. Towards a reconstruction of the pronominal systems of Proto-Cordilleran, Philippines. In Dang Liem Nguyen, ed. *South-East Asian Linguistic Studies*, vol. 3, 259–275. Canberra: Pacific Linguistics.

—— 1989. Arta, Another Philippine Negrito language. *Oceanic Linguistics* 28(1).47–74.

—— 1991. The Alta languages of the Philippines. In Ray Harlow, ed. *VICAL 2: Western Austronesian and contact languages: papers from the Fifth International Conference on Austronesian Linguistics*, 265–297. Auckland: Linguistics Society of New Zealand.

—— 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. Morphosyntactic evidence for the position of Chamorro in the Austronesian language family. In Robert Bauer, ed. *Collected papers on Southeast Asian and Pacific languages*, 63–94. Canberra: Pacific Linguistics.

Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. 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*, 521–563. Taipei: Academia Sinica.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley, Los Angeles/London: University of California Press.

Schlegel, Stuart A. 1971. *Tiruray-English lexicon*. University of California Publications in Linguistics, No. 67. Berkeley: University of California Press.

Tharp, James A. 1974. The Northern Cordilleran subgroup of Philippine languages. *Working Papers in Linguistics, Department of Linguistics, University of Hawai'i* 6(6).53–114.

Thiessen, H. Arnold. 1977. The phonemic consequences of two morphophonemic rules in Molbog. *Studies in Philippine Linguistics* 1(2).1–26.

Wolff, John U. 2007. Do the Malayo-Polynesian languages constitute a subgroup of the Austronesian language family? *Studies in Philippine Languages and Cultures* 15. 13–24.

Topping, Donal M., Pedro M. Ogo and Bernadita C. Dungca. 1975. *Chamorro-English dictionary*. PALI Language Texts: Micronesia. Honolulu: University of Hawai'i Press.

Zorc, R. David Paul. 1977. *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.

—— 1982. Where, o where, have the laryngeals gone? Austronesian laryngeals reexamined. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2, *Tracking the travellers*, 111–144. Canberra: Pacific Linguistics.

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

24 *From* ki-*N* *'get N' in Formosan languages to* ki-*V* *'get V-ed' (passive) in Rukai, Paiwan and Puyuma*

ELIZABETH ZEITOUN AND STACY F. TENG

1 Introduction[1]

The Rukai dialects exhibit morphosyntactic features that are not found in other Formosan languages. One of the most interesting, and perhaps intriguing phenomena, lies in the development of an active/passive voice dichotomy. Depending on the dialect, the active voice is marked by *w-*/*u-*/*o-* and the passive by *ki-*/*'i-*. The morpheme *ki-*/*'i-* not only occurs on verbs, but also on different types of nouns, and can be glossed as 'to obtain/get N'. The other Formosan languages, Paiwan and Puyuma in particular, are characterised by a more complex voice system. In these two languages, *ki-* is also found with the same distribution as in Rukai: *ki-* attaches to nouns, and can be glossed as 'to obtain/get N' and it attaches to verbs. When it attaches to verbs, *ki-* usually conveys a passive meaning.

The present paper accounts for the wide distribution of *ki*-N 'get N' in the Formosan languages and the more restricted development *ki*-V 'get V-ed' (passive) in Rukai,[2] Paiwan[3] and Puyuma.[4] As this morphosyntactic device is found in three contiguous languages, all spoken in Southern Taiwan, it seems necessary to first provide an overview on the genetic relationships of Rukai, Paiwan and Puyuma (§1.1) before presenting the aims of this study (§1.2).

[1] We are grateful to Robert Blust, Bethwyn Evans, Lillian M. Huang, Paul J. Li, Malcolm Ross and Laurent Sagart for comments on earlier versions of this paper. We would like to thank two anonymous reviewers for their pertinent observations. This paper was written as part of a larger project on the 'Classification and Dispersal of the Austronesians: Anthropological, Archaeological, Genetic, and Linguistic Studies Relating to Taiwan' under the direction of Paul Jen-kuei Li (NSC 94-2627-H-001-002). It is dedicated to Malcolm Ross, Elizabeth Zeitoun's long-time colleague and friend and Stacy F. Teng's PhD supervisor and mentor.

[2] Rukai includes six major dialects, Maga, Tona, Mantauran, Budai, Labuan and Tanan. Their internal relationships remain controversial. Among these, Budai is treated as the most 'conservative' and Mantauran as the most 'aberrant'. It has been demonstrated in Zeitoun (2003, 2007, in preparation) that contrary to what has been earlier assumed, Mantauran is actually genetically closer to the Labuan-Tanan-Budai cluster. The present paper draws on data from Tona.

[3] Unless otherwise indicated, the data on which this paper is based are drawn from the dialectal variant of Northern Paiwan spoken in Sandimen (Pingtung County).

[4] Puyuma is traditionally divided into two major dialects, Nanwang versus Katipul. This paper presents data from Nanwang Puyuma.

1.1 Notes on the genetic relationships of Rukai, Paiwan and Puyuma

There are fourteen extant Formosan languages, four of which are moribund. The latter are preceded with an asterisk in the list that follows: Amis, Atayal, Bunun, *Kanakanavu, Kavalan, Paiwan, *Pazeh, Puyuma, Rukai, *Saaroa, Saisiyat, Seediq, *Thao, and Tsou. A fifteenth language, Yami, is spoken on Orchid Island, which is politically part of Taiwan. Linguistically it is more closely related to the Philippine languages (Batanic subgroup).

Relationships amongst the Formosan languages are still a matter of controversy, particularly in areas where they are spoken by populations that have been in close contact for centuries, as is the case of Rukai, Paiwan and Puyuma. Various subgrouping hypotheses regarding the relationships of the Formosan languages have been advanced. Ferrell's (1969) hypothesis with a tripartite division of Formosan languages was prevalent until the 1990s. The three major subgroups included: Atayalic (Atayal and Seediq), Tsouic (Tsou, Kanakanavu and Saaroa) and Paiwanic (further divided into Paiwanic I [Rukai, Pazeh, Saisiyat, Thao, Puyuma and Paiwan] and Paiwanic II [Bunun, Siraya, Amis, Kavalan and Yami]). Since the mid-1990s, other hypotheses have emerged, based on different types of innovations: (i) phonological; (ii) morphological; and (iii) lexical. Currently, one of the most widely accepted subgrouping hypotheses, and the one we basically follow in this paper, is that of Blust (1999a). He argues, based on phonological evidence, that the Formosan languages should be classified as forming nine out of ten primary subgroups descending from Proto Austronesian (PAn) (see Figure 1). The nine groups are: Atayalic (Atayal and Seediq), East Formosan (Basay-Trobiawan, Kavalan, Amis and Siraya), Puyuma, Paiwan, Rukai, Tsouic (Tsou, Kanakanavu and Saaroa), Bunun, Western Plains (Taokas-Babuza, Papora-Hoanya, Thao) and Northwest Formosan (Saisiyat, Kulun-Pazeh). The tenth Austronesian (An) subgroup is composed of all the languages spoken outside Taiwan (Malayo-Polynesian). Blust (1999a:51) summarises the evidence regarding a possible genetic relationship between Paiwan and Puyuma as follows:

> All in all […] the evidence for a Puyumic subgroup which includes Puyuma and Paiwan is far from convincing, given the evidence for *longstanding* [our emphasis, EZ&ST] borrowing, and for now it is best to consider each of these languages a primary branch of the An family.

He does not make any comment, however, regarding the position of Rukai with respect to Paiwan and Puyuma, though he believes that this language also constitutes a primary subgroup. Four other major hypotheses have been postulated regarding the position of Rukai (as opposed to Paiwan and Puyuma) within the Formosan languages. Rukai has been viewed as: (i) subgrouping with Tsouic and forming a higher group labeled Rukai-Tsouic (Tsuchida 1976); (ii) being closer to Paiwan (Ho 1983); (iii) subgrouping with Tsouic, Paiwan, Puyuma, Amis and Bunun and forming a higher Walu-Siwaish group (Sagart 2004); and (iv) constituting the/one of the first offshoot(s) of the An family (Starosta 1994, 1995 and Ross 2009). Starosta's (1994, 1995) and Ross' (2009) views diverge in that Starosta (1994, 1995) believes that Rukai is distinct from both Tsouic and Paiwan and constitutes the first An offshoot (Starosta 1994, 1995) while Ross (2009) assumes that Rukai forms a distinct primary subgroup, as do Tsou and Puyuma in contrast to all other Austronesian languages, which form a fourth primary subgroup which he labels Nuclear Austronesian languages (see Figure 3).

1.2 Aims of the present study

The morpheme *ki-* can be glossed as 'obtain, get' as its 'core' meaning. It is prefixed to nouns (henceforth *ki*-N) and is found quite productively across the Formosan languages (see §2). *Ki-* can also be prefixed to verbs (henceforth *ki*-V), but in relatively few languages (to our knowledge, Kavalan, Rukai, Paiwan and Puyuma). The behaviour of *ki*-V is very similar in Rukai, Paiwan and Puyuma. In these three languages, *ki-* functions most notably as a passive (see §3).

The goals of the present paper are as follows:

(i) to describe the distribution and function of the prefix *ki-* when followed by a noun in the Formosan languages and show that it is a viable candidate for reconstruction at the PAn level (§2).

(ii) to describe and compare the distribution and functions of *ki-* when followed by a verb in Rukai, Paiwan and Puyuma (§3).

(iii) to determine whether *ki*-V represents borrowing, shared innovation or parallel development in these three languages and to discuss the grammaticalisation path of *ki-* (§4).

2 *ki*-N 'obtain-N' in Formosan languages

As shown in (1)–(6) below, the prefix *ki-* 'obtain, get' followed by a noun is found productively in Saisiyat, Kavalan, Kanakanavu, Saaroa, Rukai, Puyuma and Paiwan. The derived form, *ki*-N, functions as verb in all these languages.

(1) TUNGHO SAISIYAT[5]

a.	*kaehoey*	'tree, wood, brushwood'	*ki-kaehoey*	'gather, chop brushwood'
b.	*pongaeh*	'flower'	*ki-pongaeh*	'pick flowers'

(2) KAVALAN (PAn *k > Kav *q*; Li and Tsuchida 2006:229, 441)[6]

a.	*paRin*	'tree, wood, brushwood'	*qi-paRin*	'gather firewood'
b.	*tamun*	'vegetable'	*qi-tamun*	'pick vegetables'

(3) a. KANAKANAVU (Ho 1997:240)

tamemi	'sweet potato'	*ki-tamemi*	'gather sweet potatoes'

b. SAAROA (Li pers. comm)

mairange	'sweet potato'	*ki-mairange*	'gather sweet potatoes'

[5] The orthographic symbols employed in this paper include: e [ə] ae [æ], oe [œ], ' [ʔ], tr [ʈ], dr [ɖ], lr [ɭ], dh [ð], sh [ʃ], lh [ɬ], th [θ], tj [c], dj [ɟ], ng [ŋ] and r [R].There is no attempt to distinguish between prefixes and clitics. Abbreviations are: AV - Actor Voice; CAUS - Causative; CV - Circumstantial Voice; DYN - Dynamic; EXCL - Exclusive; FIN - Finite; GEN - Genitive; INCL - Inclusive; INDF - Indefinite; IRR - Irrealis; Kav - Kavalan; LIG - Ligature; LV - Locative Voice; NAGTPASS - Non-agentive Passive; NAV - Non Actor Voice; NEG - Negation; NFIN - Non Finite; NOM - Nominative; OBL - Oblique; Pai - Paiwan; PASS - Passive; PERF - Perfective; PL - Plural; PROJ - Projective; PV - Patient Voice; REAL - Realis; RED - Reduplication; STAT - Stative; SUBJ - Subjunctive; TOP - Topic.

[6] Both Tungho Saisiyat and Kavalan exhibit at least one other prefix with a similar meaning. In Tungho Saisiyat, *ki*-N means 'to gather, harvest (with an instrument)' as opposed to *ti*-N 'to gather, harvest (with one's hands). Kavalan has another prefix that means 'catch, get', cf. *Ri-* as in *Ri-baut* 'to fish' < *baut* 'fish', *Ri-alam* 'catch birds' < *alam* 'bird', *Ri-krisiw* 'get paid' < *krisiw* 'money', *Ri-szang* 'expose to the sun' < *szang* 'sun' (Li and Tsuchida 2006:19).

In Tona Rukai, Nanwang Puyuma and Northern Paiwan, *ki-* appears on a variety of nouns and means ‘gather/harvest, fetch, get, hunt/kill etc.’, as shown in (4)–(6) respectively.

(4) TONA RUKAI

a.	*becenge*	‘millet’	*ky-a-becenge*	‘harvest millet’
b.	*pagay*	‘rice’	*ky-a-pagay*	‘harvest rice’
c.	*’angato*	‘brushwood, tree’	*ky-a-’angato*	‘gather brushwood’
d.	*enay*	‘water’	*ky-a-enay*	‘fetch water’
e.	*paiso*	‘money’	*ky-a-paiso*	‘earn money’
f.	*baa*	‘enemy’	*ky-a-baa*	‘kill (an) enemy’
g.	*comay*	‘bear’	*ky-a-comay*	‘kill (a) bear’

(5) NANWANG PUYUMA (Teng 2008 and Cauquelin 1991)

a.	*’aputr*	‘flower’	*ki-’aputr*	‘pick flowers’
b.	*kawi*	‘wood’	*ki-kawi*	‘chop wood’
c.	*asepan*	‘sugarcane’	*ki-asepan*	‘hack sugarcane’
d.	*kuraw*	‘fish’	*ki-kuraw*	‘fish’
e.	*paisu*	‘money’	*ki-paisu*	‘earn money’
f.	*ni’en*	‘neck’	*ki-ni’en*	‘behead’
f’.	*tranguru’*	‘head’	*ki-tranguru’*	‘behead’
g.	*’ala*	‘enemy’	*ki-’ala*	‘kill an enemy’
h.	*babuy*	‘wild boar’	*ki-babuy*	‘kill (a) wild boar’

(6) NORTHERN PAIWAN

a.	*va’u*	‘millet’	*ki-va’u*	‘harvest millet’
b.	*paday*	‘rice’	*ki-paday*	‘harvest rice’
c.	*kasiw*	‘wood, brushwood’	*ki-kasiw*	‘gather brushwood’
d.	*tevus*	‘sugarcane’	*ki-tevus*	‘hack sugarcane’
e.	*zalum*	‘water’	*ki-zalum*	‘fetch water’
f.	*paisu*	‘money’	*ki-paisu*	‘get/earn money’
g.	*’ulu*	‘head’	*ki-’ulu*	‘behead’
h.	*vavuy*	‘wild boar’	*ki-vavuy*	‘kill (a) wild boar’

It is reasonable to believe that the prefix *ki-* has undergone a semantic shift. In Isbukun Bunun and in Tungho Saisiyat, *ki*-N means ‘remove from’ but seems to be used specifically with nouns related to body parts, as shown in (7) and (8).

(7) ISBUKUN BUNUN

a.	*ngisngis*	‘beard’	*ki-ngisngis*	‘shave’ (Lin et al. 2001:88)
b.	*saip*	‘body hair’	*ki-saip*	‘remove body hair from the face’ (ibid. p.42)

(8) TUNGHO SAISIYAT

nipen	‘tooth’	*ki-nipen*	‘pull out a tooth’ (Yeh 2003:44)

In Thao, there seems *a priori* to be no prefix *ki-* meaning ‘harvest, gather’ but there is a prefix, cf. *kin-*, with such a meaning. It has two other allomorphs *kim-* (before bilabials) and *kig-* (before velars).[7]

[7] Kaufman (2009) proposes that the *m-* (~*n-*~*g-*) prefix refers to a plural.

(9) THAO (Blust 2003:104)

a.	*bailu*	'beans'	*kim-bailu*	'pick beans'
b.	*bukay*	'flower'	*kim-bukay*	'pick flowers'
c.	*fatu*	'stone'	*kin-fatu*	'gather stones'
d.	*lapat*	'guavas'	*kin-lapat*	'pick guavas'
e.	*lhuzush*	'plums'	*kin-lhuzush*	'gather plums'
f.	*rusaw*	'fish'	*kin-rusaw*	'catch fish, fish'
g.	*kucun*	'shrimp'	*kig-kucun*	'collect shrimps'
h.	*qati*	'bamboo shoots'	*kig-qati*	'gather bamboo shoots'

To date, *ki*-N has not been reported in Pazeh (Blust 1999b; Li and Tsuchida 2001), Atayal (Egerod 1999), Seediq (Pecoraro 1979) or Amis (Fey 1986).[8]

A summary of the data provided in this section is given in Table 1.

Table 1: Distribution of the prefix *ki*-N in the Formosan languages

Language	Form	Meaning	
		1. 'to get, harvest'	2. 'to remove'
Kavalan	*qi-*	+	–
Kanakanavu	*ki-*	+	–
Saaroa	*ki-*	+	–
Rukai	*ki-*	+	–
Puyuma	*ki-*	+	–
Paiwan	*ki-*	+	–
Saisiyat	*ki-*	+	+
Bunun	*ki-*	–	+
Thao	*(kin-)*	+	–
Pazeh	–	–	–
Atayal	–	–	–
Seediq	–	–	–
Amis	–	–	–

A conclusion imposes itself at this point: whether we follow Blust's (1999a), Sagart's (2004) or Ross' (2009) classification of Austronesian languages, the distribution of *ki*-N in the Formosan languages shows that it should be reconstructed for PAn, as demonstrated by Figures 1, 2 and 3 where the distribution of *ki-* across the Formosan languages (bolded) can be seen.

[8] In Nataoran Amis, there is a prefix *li-* meaning 'to get, to obtain' as shown as follows:

li-tinaiq	'get/take instestines'	<	*tinaiq*	'instestines'
mi-li-pida	'get a salary'	<	*pida*	'money'
mi-li-panay	'harvest rice'	<	*panay*	'rice'
mi-li-tangal	'kill (i.e. go headhunting)'	<	*tangal*	'head'

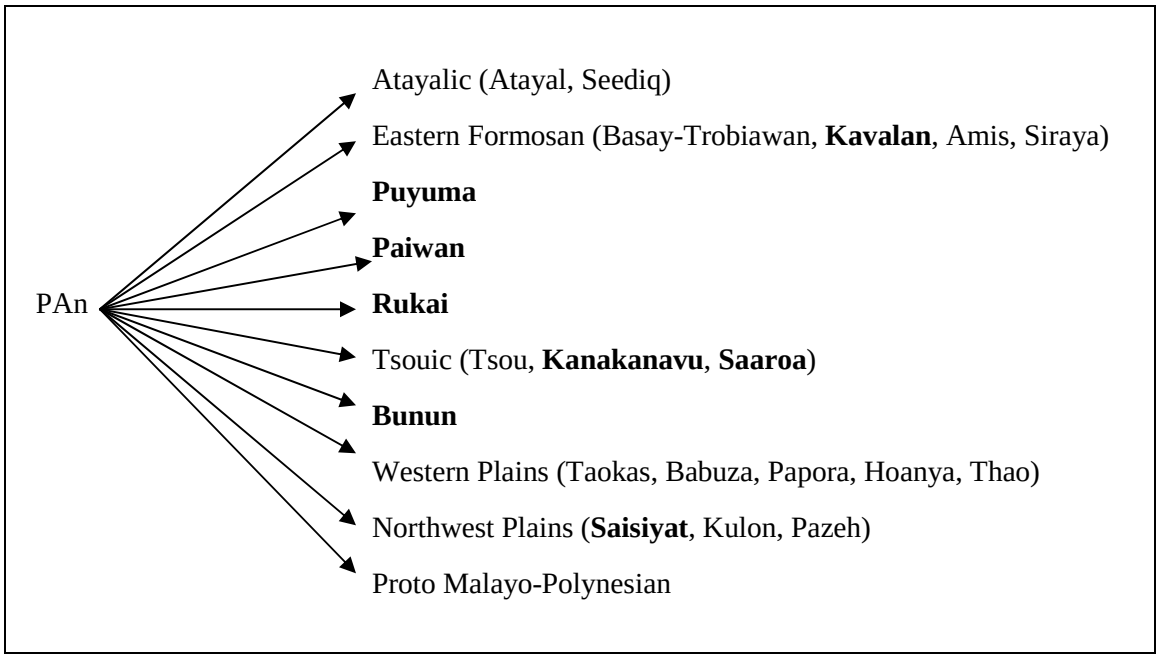


Figure 1: Blust's (1999:45) subgrouping hypothesis

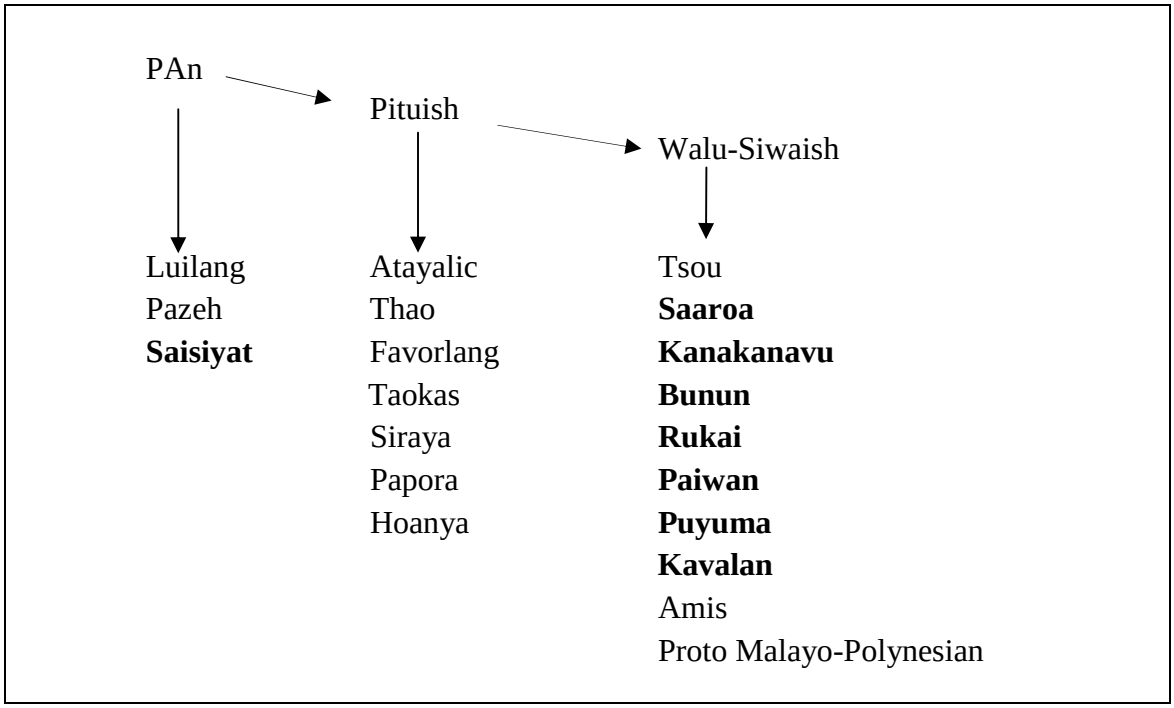


Figure 2: Sagart's (2004:421) subgrouping hypothesis

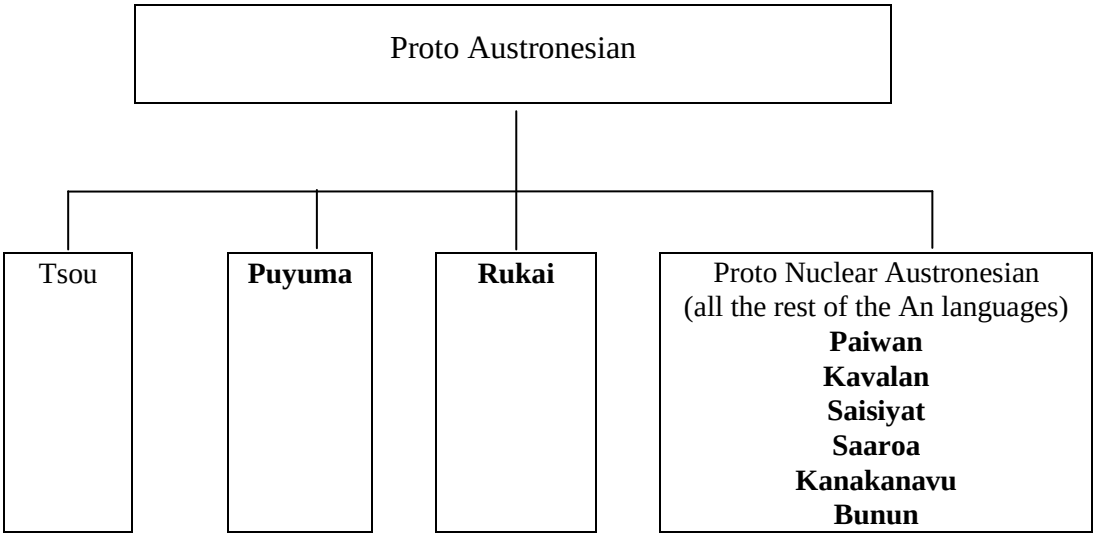


Figure 3: Ross' (2009) subgrouping hypothesis

3 Distribution and functions of *ki*-V in Rukai, Paiwan and Puyuma

The major distinction between Rukai, Puyuma and Paiwan lies in their voice system. Rukai displays an active/passive voice dichotomy while the other Formosan languages — Puyuma and Paiwan in particular — are characterised by a more complex voice system. Two major voices are found in these two languages: A(ctor) V(oice) and N(on) A(ctor) V(oice),[9] which further includes the following subtypes: P(atient) V(oice), L(ocative) V(oice), and C(onveyance) V(oice).[10] An overview of the voice distinctions in Rukai, Puyuma and Paiwan is further given in §3.1.

In these three languages, *ki-* followed by a verb stem has two different (though related) meanings: (i) it refers to a passive, and (ii) it indicates a reflexive.[11] The distribution and

---

9 In Puyuma, AV clauses are analysed as intransitive as opposed to NAV clauses (for a detailed discussion on this matter, see Ross and Teng 2005 and Teng 2008).

10 As demonstrated in earlier papers (cf. Zeitoun et al. 1996; Zeitoun and Huang 1997), voice (or focus) interacts closely with mood and aspect. In the present section, we provide examples in the realis mood only.

11 The use of the same marker as a passive or a reflexive has been widely discussed in general linguistics. Shibatani (1985:825) shows that that '[i]n several Indo-European and American Indian languages, the passive/reflexive/reciprocal correlation is well known. […] the same morphology is employed for all three constructions, or for a passive and one other construction' and further argues 'that the passive/ reflexive/reciprocal correlation arises largely from a semantic property of these constructions: in all of them, surface subjects are affected. In the passive, the subject is affected by an external agent; in the reflexive, by itself; in the reciprocal, by the partner.'

It is instructive to note that:

(a) *ki-* 'self' has only been reported in two Rukai dialects, Budai and Tanan (see examples in (i); Li 1975:260), but in these two dialects, *ki-* competes with another reflexive form, cf. Budai *ngi-*, which occurs much more productively, e.g. Budai *ngy-a-pa-pa-pacay* 'kill oneself' (< *ngi-* 'self', *a-* 'realis', *pa-* 'reduplication', *pa-* 'Caus', *pacay* 'die');

(i) BUDAI RUKAI

a.	*vavavange*	'play'	*ky-a-vavavange*	'play by oneself'
b.	*papacay*	'kill'	*ky-a-pa-papacay*	'kill oneself'

(b) in Paiwan, *ki-* has been shown to exhibit an array of functions (cf. Ferrell 1982 and Chang 2006), including 'reflexive', as in (ii); and

functions of *ki*-V as passive are further discussed in §3.2. *Ki*-V clauses will be compared, on the one hand, to *ki*-N clauses (§3.3), and to NAV clauses on the other (§3.4). The reflexive function of *ki-* will not be discussed in this paper, as the complexity of the linguistic data observed needs to be further investigated.

3.1 Overview of the voice distinctions in Rukai, Puyuma and Paiwan

In Tona Rukai, dynamic verbs are marked by *w-* in the active voice (realis), as in (10a). Other prefixes might also be used, but far less productively. Stative verbs are commonly marked by *ma-*, as in (10b).

(10) TONA RUKAI

a. *a-nakay soa'a ka w-a-ka'ace na atho.*
TOP-that snake TOP DYN.FIN-REAL-bite OBL dog
'That snake bit a dog'

b. *ma-boti'i na ngiaw.*
STAT.FIN-blind NOM cat
'The cat is blind.'

There are two passive forms in Tona Rukai. The first is agentive, as in (11a) and the second is agentless, as in (11b). The agentive passive is realised as *ki-*,[12] the agentless passive as *'i-*.

(11) TONA RUKAI

a. *a-nakay atho ka ky-a-ka'ace na soa'a.*
TOP-that dog TOP PASS-REAL-bite OBL snake
'That dog was bitten by a snake.'

b. *'y-a-silape-nga makasi i-okono kiname.*
NAGTPASS-REAL-look for-already but NEG-DYN.NFIN:find 1PL.EXCL.NOM
lit: 'It was looked for but we did not find it'

Two major voices are found in Puyuma: AV and NAV. The latter has three subtypes: PV *-aw*, LV *-ay*, and CV *-anay*, as shown in (12):

(12) NANWANG PUYUMA

a. *tusuk-ku dra lrutung.*
<AV>pierce-1SG.NOM INDF.OBL monkey
'I speared a monkey.' (Ross and Teng 2005:749)

(ii) NORTHERN PAIWAN

a.	*pacay*	'die'	*ki-pacay*	'commit suicide'
b.	*se'as*	'chop'	*ki-se'as*	'chop oneself'
c.	*mavanaw*	'bathe'	*ki-pavanaw*	'bathe oneself'
c.	*ngua'ngua'*	'beautiful'	*ki-lengua'*	'dress up beautifully, make up'

(c) in Puyuma, *ki*-V has been glossed as a reflexive by Cauquelin (1991:20), cf. *natay* 'die' ~ *ki-natay* 'commit suicide' but our own informants reject such a construction.

[12] *ki-* is realised as *ky-*, and *'i-* as *'y-* when followed by the realis marker *a-*.

b. *ku-tusuk-aw* *na* *lrutung.*
1SG.GEN-pierce-PV NOM monkey
‘I speared the monkey.’

c. *ku-tusuk-ay* *dra* *da’um* *nantu* *tranguru’* *kana* *lrutung.*
1SG.GEN-pierce-LV INDF.OBL needle NOM:3P head DEF.OBL monkey
‘I pierced the monkey’s head with a needle.’

d. *ku-tusuk-anay* *na* *derederan* *dra* *lrutung.*
1SG.GEN-pierce-CV NOM spear type INDF.OBL monkey
‘I speared monkeys with the *derederan* (-kind of spear).’

Two major voices are also found in Paiwan: AV versus NAV, which further divides into: PV *<in>/-en*, LV *-an*, and CV *si-*, as shown in (13). Stative verbs are either unmarked or marked by *ma-*, as illustrated in (14).

(13) NORTHERN PAIWAN

a. *tekelr* *ta* *vava* *ti* *kui.*
drink<AV>drink OBL wine NOM Kui
‘Kui drinks/drank wine.’

b. *t<in>ekelr-anga* *a* *vava* *ni* *kui.*
drink<PV:PERF>drink-already NOM wine GEN Kui
‘Kui has already drunk the wine.’

c. *pa-tekelr-an* *ti* *su-kaka* *tuazua* *zalum.*
CAUS-drink-LV NOM 2SG.GEN-elder sibling OBL:that water
‘That (glass of) water is for your elder brother/sister to drink.’

d. *si-tekelr* *ta* *cemelr.*
CV-drink OBL grass
‘(This glass) is used to drink medicine.’

(14) a. *Ø-tengelray* *ti* *kui* *tai* *muni.*
STAT.FIN-love NOM Kui OBL Muni
‘Kui loves Muni.’

b. *ma-lreva* *ti* *’umi.*
STAT.FIN-happy NOM Umi
‘Umi is happy.’

3.2 Notes on the distribution and functions of *ki*-V in Rukai, Puyuma and Paiwan

In Rukai, the passive prefix *ki-* can occur with dynamic and stative verbs, as shown in (15). Only inherently transitive verbs can be passivised. Note that although passive forms can be easily elicited in the Rukai dialects, and in Tona in particular, they are not frequently found in texts.

(15) TONA RUKAI

(i) *ki-* ‘(agentive) passive’ in co-occurrence with (transitive) dynamic verbs

a.	*w-a-aba*	‘carry on back’	*ky-a-aba*	‘be carried on back’
b.	*w-a-do’o*	‘cook’	*ky-a-do’o*	‘be cooked’
c.	*w-a-elebe*	‘close’	*ky-a-elebe*	‘be closed’

d.	*w-a-igo'o*	'know'	*ky-a-igo'o*	'be known, be famous'
e.	*w-a-ka'ace*	'bite'	*ky-a-ka'ace*	'be bitten'
f.	*w-a-lapo*	'raise (an animal)'	*ky-a-lapo*	'be raised'
g.	*w-a-pii*	'choose'	*ky-a-pii*	'be chosen'
h.	*w-a-pwalra*	'catch'	*ky-a-pwalra*	'be caught'
i.	*w-a-salaa'a*	'chase'	*ky-a-salaa'a*	'be chased'
j.	*w-a-sititi*	'beat'	*ky-a-sititi*	'be beaten'
k.	*w-a-tobi*	'cry'	*ky-a-tobi*	'beg'

(ii) *ki-* '(agentive) passive' in co-occurrence with (transitive) stative verbs

a.	*ma-dalame*	'like, love'	*ky-a-ka-dalame*	'be liked, loved'
b.	*ma-ga'aoco*	'scold'	*ky-a-ka-ga'aoco*	'be scolded'

As in Tona Rukai, *ki-* can attach to different types of verbs in Puyuma and in Paiwan.

(16) NANWANG PUYUMA (Teng 2008; Cauquelin 1991)

(i) *ki-* in cooccurrence with dynamic verbs

a.	*abak*	'fill in'	*ki-abak*	'be filled'
b.	*adras*	'lift'	*ki-adras*	'be lifted'
c.	*alrak*	'take'	*ki-alrak*	'be taken away'
d.	*asal*	'move'	*ki-asal*	'be moved'
e.	*atrab*	'cover'	*ki-atrab*	'be covered'
f.	*baluk*	'wake'	*ki-baluk*	'be waken up'
g.	*bekas*	'attack'	*ki-bekas*	'be attacked'
h.	*beray*	'give'	*ki-beray*	'be given, ask for'
i.	*bulras*	'exchange'	*ki-bulras*	'borrow'
j.	*da'ul*	'inform'	*ki-da'ul*	'be informed'
k.	*dirus*	'wash'	*ki-dirus*	'be washed'
l.	*drimutr*	'catch'	*ki-drimutr*	'be caught'
m.	*kasu*	'bring'	*ki-kasu*	'be brought'
n.	*lrelrep*	'catch up'	*ki-lrelrep*	'be caught up'
o.	*na'u*	'see'	*ki-na'u*	'be seen, looked after'
p.	*padek*	'carry'	*ki-padek*	'be carried'
q.	*pilang*	'lead'	*ki-pilang*	'be led'
r.	*retra'*	'put down'	*ki-retra'*	'be discarded'
s.	*sabana*	'cheat'	*ki-sabana*	'be cheated'
t.	*salraw*	'surpass'	*ki-salraw*	'be surpassed'
u.	*tarama*	'bully'	*ki-tarama*	'be bullied'
v.	*tenges*	'tie up'	*ki-tenges*	'be tied up'
w.	*tuludr*	'put s.t in s.o's hands'	*ki-tuludr*	'be put in one's hands, accept, receive'
x.	*trakaw*	'steal'	*ki-trakaw*	'be stolen'
y.	*trangis*	'cry'	*ki-trangis*	'beg'

(ii) *ki-* in cooccurrence with stative verbs

a.	*bangabang*	'be busy'	*ki-bangabang*	'kill time'
b.	*laman*	'pity'	*ki-laman*	'be pitied, receive pity from others'

c.	*litek*	'be cold'	*ki-litek*	'catch a cold'
d.	*sagar*	'like'	*ki-sagar*	'be liked, receive love from others'

(17) NORTHERN PAIWAN

(i) *ki-* in cooccurrence with dynamic verbs

a.	*pangulr*	'beat'	*ki-pangulr*	'get beaten'
b.	*zurung*	'push (down)'	*ki-zurung*	'get pushed (down)'
c.	*pasedjan*	'lend'	*ki-pasedjan*	'borrow'
d.	*'aung*	'cry'	*ki-'aung*	'beg'

(ii) *ki-* in cooccurrence with stative verbs

madodo	'scold'	*ki-kadodo*	'be scolded'

3.3 Major distinctions between *ki*-N clauses and *ki*-V clauses in Rukai, Puyuma and Paiwan

There are major distinctions between *ki*-N clauses and *ki*-V clauses in Rukai, Puyuma and Paiwan. They include the marking of nominal arguments (§3.3.1) and the co-occurrence of *ki-* with different verbal affixes (§3.3.2).

3.3.1 Marking of arguments

In the three languages, the *agent* is the nominative argument in *ki*-N clauses, as shown in (18).

(18) *ki*-N clause-type

a. TONA RUKAI

… ky-a-cila-cilay kake silape na acilay
get-REAL-RED-water 1S.NOM DYN.SUBJ:look.for OBL water

m-wa 'ongolo.
DYN.SUBJ-go DYN.SUBJ:drink
'… I would fetch water to drink.' (FLA DRUTo_11_010_b)[13]

b. NANWANG PUYUMA

ki-'aputr-ku-la.
get-flower-1SG.NOM-PERF
'I have picked flowers.'

c. NORTHERN PAIWAN

na-ki-va'u-aken.
PERF-get-millet-1SG.NOM
'I harvested millet.'

In *ki*-+V clauses, on the other hand, the nominative argument is the *patient* and the agent is marked as oblique, as illustrated in (19).

[13] Example from the Formosan Language Archive (FLA), see http://formosan.sinica.edu.tw.

(19) *ki*-V clause-type

a. TONA RUKAI
ki-cengele kake ’osam-ane m-ya.
PASS-see 1SG.NOM king-OBL SUBJ.DYN-so
‘I said: “I was seen by the king.”’ (FLA DRUTo_08_006_b)

b. NANWANG PUYUMA
ki-drimutr i senayan kana kinsas.
PASS-seize NOM Senayan OBL policeman
‘Senayan was seized by the policeman.’

c. NORTHERN PAIWAN
ki-zurung-aken tay kina.
PASS-push-1SG.NOM OBL mother
‘I was pushed by mother.’

The contrast in the marking of nominal arguments in *ki*-N and *ki*-V clauses is tabulated as follows:

Table 2: Contrast in the marking of nominal arguments in *ki*-N and *ki*-V clauses

	ki-N		*ki*-V		
	No. of arguments	Marking of AGENT	No. of arguments	Marking of arguments AGENT	PATIENT
Rukai Puyuma Paiwan	1	NOM	2	OBL	NOM

3.3.2 Co-occurrence of *ki*-N and *ki*-V with different verbal affixes

In Rukai, Paiwan and Puyuma, denominal verbs like those in (20)–(22) do not occur with any voice affixes. In other words, *ki*-N forms belong to the class of verbs in which the active voice (in Tona Rukai)/AV (in Puyuma and Paiwan) is zero-marked.

(20) TONA RUKAI

a.	*ki-becenge*	‘harvest millet’	**w-a-ki-becenge*	**ky-a-ki-becenge*
b.	*ki-’angato*	‘gather brushwood’	**w-a-ki-’angato*	**ky-a-ki-’angato*
c.	*ki-paiso*	‘earn money’	**w-a-ki-paiso*	**ky-a-ki-paiso*

(21) NANWANG PUYUMA

a.	*ki-’aputr*	‘pick flowers’	**ki-’aputr*	**ki-’aputr-aw*
b.	*ki-kawi*	‘hack woods’	**ki-kawi*	**ki-kawi-aw*
c.	*ki-asepan*	‘hack sugarcane’	**ki-asepan*	**ki-asepan-aw*

(22) NORTHERN PAIWAN

a.	*ki-va’u*	‘get or harvest millet’	**ki-va’u*	**ki-va’u-en*
b.	*ki-kasiw*	‘chop woods’	**ki-kasiw*	**ki-kasiw-en*
c.	*ki-tevus*	‘hack sugarcane’	**ki-tevus*	**ki-tevus-en*

One major distinction between Puyuma and Paiwan is that Puyuma distinguishes three sets of *ki*-derived verbs. While denominal verbs like those in (21) do not occur with any voice affixes, certain denominal verbs, like *ki-lengaw* can appear with NAV voice affixes, but cannot occur with the AV <*em*> (23a-b), while yet other denominal verbs like *anger* 'thought', *rami* 'root' can take both AV and NAV voice affixes, as shown in (24a-b).

(23) NANWANG PUYUMA (from Teng 2008)

a. *tu-ki-lengaw-ay i tinataw.*
3GEN-get-sound-LV SG.NOM his mother
'He listened to his mother.'

b. **ki-lengaw i senayan kana sinsi.*
get<AV>-sound SG.NOM Senayan OBL teacher

(24) a. *an ki-anger-ta i, ...*
when get<AV>-thought-1PL.INCL.NOM TOP
'When we thought about it …'

b. *tu-ki-anger-aw tu-pi-amanay-an*
3GEN-get-thought-PV his.NOM-have-what-NMZ
'He thought about what he had.'

Note that in Nanwang Puyuma, when *ki-* is followed by a verb, it does not occur with any voice affixes, as shown in (25). This constraint is also observed in Paiwan, cf. (26).

(25) NANWANG PUYUMA

a. **ki-drimutr i senayan kana kinsas.*
PASS<AV>seize NOM Senayan OBL policeman

b. **tu-ki-drimutr-aw i senayan kana kinsas.*
3GEN-PASS-seize-PV NOM Senayan OBL policeman

(26) NORTHERN PAIWAN

a. **ki-zurung-aken tay kina.*
PASS<AV>-push-1SG.NOM OBL mother

b. **k<in>i-zurung-aken ni kina.*
PASS<PV.PERF>-push-1SG.NOM GEN mother

In the three languages, *ki*-V can occur with other verbal affixes. For example, *si-* 'raise', *pa-* 'Caus' in Tona (27), *pa-* 'Caus-' and *pu-* 'CausMvt' in Puyuma (28), *pa-* 'Caus', *pu-* 'CausMvt' and *su-* 'remove' in Paiwan (29).

(27) TONA RUKAI

a.	*si-valake*	'raise (a child)'	*ky-a-si-valake*	'be raised'
b.	*pa-'ongolo*	'make…drink'	*ky-a-pa-'ongolo*	'ask for a drink'
c.	*pa-kane*	'feed'	*ky-a-pa-kane*	'ask for food'
d.	*pa-dakili*	'make…kneel down'	*ky-a-pa-dakili*	'be asked to kneel down'

(28) NANWANG PUYUMA

a.	*pa-takesi*	'educate, teach'	*ki-pa-takesi*	'be taught, receive (an) education'
b.	*pa-trekelr*	'make…drink'	*ki-pa-trekelr*	'ask for a drink'
c.	*pa-kan*	'feed'	*ki-pa-kan*	'be fed'

d.	*pu-ngalradr*	'give a name'	*ki-pu-ngalradr*	'ask s.o to give a name'
e.	*pu-walak*	'make pregnant'	*ki-pu-walak*	'get pregnant'

(29) NORTHERN PAIWAN

a.	*pa-kim*	'make…look for'	*ki-pa-kim*	'be looked for'
b.	*pa-tekelr*	'make…drink'	*ki-pa-tekelr*	'ask for a drink'
c.	*pa-kan*	'feed'	*ki-pa-kan*	'ask for food'
d.	*pu-cemelr*	'put grass on'	*ki-pu-cemelr*	'have grass put on, get cured'
e.	*su-alis*	'pull out tooth'	*ki-su-alis*	'have a tooth removed'

Table 3: Co-occurrence of *ki*-N with verbal affixes

Language	*ki*<AV>N	*ki*-N-NAV
Tona Rukai	–	–
Nanwang Puyuma		
kawi 'wood' etc.	–	–
lengaw 'sound'	–	+
anger 'thought'	+	+
Northern Paiwan	–	–

Table 4: Co-occurrence of *ki*-V with verbal affixes

Language	*ki*<AV>V	*ki*-V-NAV	*ki-pa*-V	*ki-pu*-V
Tona Rukai	–	–	+	–
Nanwang Puyuma	–	–	+	+
Northern Paiwan	–	–	+	+

3.4 Major distinctions between *ki*-V clauses and NAV clauses in Puyuma and Paiwan

In Paiwan and Puyuma, NAV clauses are distinguished from *ki*-passive clauses syntactically and semantically. As shown below, on the syntactic level, *ki*-passive clauses differ from NAV clauses with respect to the case marking of the agent (see examples (30)–(31)) and whether it can function as a shared argument or not (see examples (32)–(33)).

In Nanwang Puyuma, the patient, *Senayan* in (30), is marked nominative both in *ki*-passive clauses and in NAV clauses, as shown in (30a-b). The agent in these clauses, *kinsas* 'policeman', is marked oblique in both constructions, but in a NAV clause the oblique-marked agent is always cross-referenced on the verb with a genitive pronoun, as shown in (30b-b').

(30) NANWANG PUYUMA

a. *ki-drimutr i senayan kana kinsas.*
PASS-seize NOM Senayan OBL policeman
'Senayan was seized by the policeman.'

b. *tu-drimutr-aw i senayan kana kinsas.*
3GEN-seize-PV NOM Senayan OBL policeman
"The policeman seized Senayan."

b'. **Ø-drimutr-aw i senayan kana kinsas.*
Ø-seize-PV NOM Senayan OBL policeman

In Northern Paiwan, the agent is marked as genitive in NAV clauses as in (31a), but is marked as oblique in *ki*-passive clauses, as in (31b):

(31) NORTHERN PAIWAN

a. *d<in>ame' ti kai nua kisacu.*
seize<PV>seize NOM Kai GEN policeman.
'Kai was seized by the policeman.'

b. *ki-dame' ti kai tua kisacu.*
PASS-seize NOM Kai OBL policeman
'Kai was seized by the policeman.'

The agent of a NAV clause can function as the shared argument of each verb in a serial verb construction, i.e. the argument expressed as agent of the first verb and co-referential with the deleted subject of the second verb. The agent of a *ki*-clause cannot be such a shared argument. In (32a), the agent *sinsi* 'teacher' is the shared argument of *drimutr* 'to serve' and *pa-karun* 'to make work'; (32b) with *ki-*, on the other hand, is not acceptable. The same restriction is found in Northern Paiwan, as illustrated in (33a-b).

(32) NANWANG PUYUMA

a. *tu-drimutr-aw kana sinsi pa-karun.*
3GEN-seize-PV OBL teacher CAUS-work
'The teacher seized him to make him work.'

b. **ki-drimutr kana sinsi pa-karun.*
PASS-seize OBL teacher CAUS-work

(33) NORTHERN PAIWAN

a. *s<in>ekaul a pa-ka-sengseng nua sinsi ti kai.*
send<PV.PERF>send LIG CAUS-STAT-work GEN teacher NOM Kai
'The teacher sent Kai to work.'

b. **ki-sekaul a pa-ka-sengseng tua sinsi ti kai.*
PASS-send LIG CAUS-STAT-work OBL teacher NOM Kai

In addition to the syntactic differences discussed above, on the semantic level, the patient of these two constructions exhibits different degrees of volition. The patient of a *ki*-passive clause is highly volitional and strongly intends the action to be carried out, while there is no such implication in a NAV clause, as illustrated by (34) and (35). Furthermore, a *ki*-passive verb can appear after verbs denoting strong intention/desire, such as 'go to' or 'want' and form a serial verb construction, while a NAV verb cannot appear in the same position.[14]

[14] The ungrammaticality of (34c) and (35c) results from the restriction imposed on serial verb constructions, whereby non-initial verbs must be marked as AV/intransitive. In this sense, *ki*-marked passive verbs seem to be morphologically intransitive.

(34) NANWANG PUYUMA

a. *a-uka-ku* *ki-tusuk-a.*
IRR-go-1SG.NOM PASS-pierce-PROJ
'I am going to get vaccinated.' (Cauquelin 1991:216)

b. *ku-tusuk-aw* *na* *lrutung.*
1SG.GEN-pierce-PV NOM monkey
'I speared the monkey.' (Ross and Teng 2005:749)

c. **a-uka-ku* *tusuk-aw* *na* *lrutung*.
IRR-go-1S.NOM pierce-PV NOM monkey

(35) NORTHERN PAIWAN

a. *vaik-aken* *a* *ki-pangulr.*
go-1S.NOM LIG PASS-beat
'I went to get beaten.'

b. *ku-p<in>angulr* *ti* *kivi.*
1SG.GEN-beat<PV.PERF>beat NOM Kivi
'I beat Kivi.'

c. **vaik-aken* *p<in>angulr* *ti* *kivi.*
go-1S.NOM beat<PV.PERF>beat NOM Kivi

Note that the volition of the nominative argument in *ki*-passive clauses is higher in Puyuma and Paiwan than in Rukai, where it is unmarked.

Table 5: Contrast between *ki*-clauses and NAV clauses in Puyuma and Paiwan

	***ki*-clauses**				**NAV-clauses**			
	Argument marking		Shared argument	Volition of PAT	Argument marking		Shared argument	Volition of PAT
	PAT	AGT			PAT	AGT		
Nanwan Puyuma	NOM	OBL	no	higher	NOM	OBL + GEN prn	yes	lower
Northern Paiwan	NOM	OBL	no	higher	NOM	GEN	yes	lower

4 *ki*-V in Rukai, Paiwan and Puyuma: borrowing, shared innovation or parallel development?

The fact that *ki*-V 'passive' occurs in three contiguous languages, namely Rukai, Paiwan and Puyuma, leads to a question regarding its origin: does this use of *ki*- represent a shared innovation, a loan, or a parallel development?

Besides the fact that in many subgrouping hypotheses (cf. §1.1), there is no other evidence that Rukai, Paiwan and Puyuma form a subgroup, there are two other reasons to reject the hypothesis that *ki*-V 'passive' represents an exclusively shared innovation in these three languages.

First, *ki*-N 'obtain/get-N' is found in several Formosan languages and apparently already existed in PAn. It is thus not exclusively found in Rukai, Paiwan and Puyuma.

Second, the grammaticalisation to *ki*-N to *ki*-V represents a natural development cross-linguistically, as shown most notably in Heine and Kuteva (2002:145–146). To give but one example, in Rodrigues Creole (French-based), *gaŷ* (cf. French 'gagner') is used as 'get' in (36a) but as a passive in (36b).

(36) RODRIGUES CREOLE (Corne 1977:164–165, quoted from Heine and Kuteva 2002:146)

a. *mo fin gaŷ sa avekli.*
1SG CPL get it with3SG
'I got it from him.'

b. *lisiẽ i gaŷ morde ek pis.*
dog 3SG get bite with flea
'Dogs get bitten by fleas.'

The first hypothesis having been rejected, any similarities among these three languages must be the result of diffusion through contact or parallel development, and neither of these two possibilities can be so easily dismissed. We discuss these two hypotheses in turn.

4.1 Diffusion through contact

Diffusion through contact implies the transfer of certain linguistic traits from one language to another. Heine and Kuteva (2005:4) state that

> transfer tends to be based on […] some way of equating a grammatical concept or structure Mx of language M (= the model language) with a grammatical concept or structure Rx of language R (= the replica language).

While this hypothesis is appealing, there are at least two reasons for not endorsing it completely.

First, to posit the transfer of *ki*-V from one language to another would require that we first identify the language M and the two languages R. This seems difficult at this point, as it is well known that through their history there has been evidence for *longstanding* [Blust's (1999a) term] borrowing between Puyuma and Paiwan, Paiwan and (Budai) Rukai and (Tanan) Rukai and Puyuma.

Second, it is also well-known from published material that *ki*-V has different meanings in these three languages, namely 'intention', 'passive' and 'reflexive' in Paiwan (Ferrell 1982), 'passive' and 'reflexive' in Rukai, and 'passive', 'change of direction', 'middle' and perhaps 'reflexive' in Puyuma (Teng 2008). In other words, *ki*-V has evolved differently in these three languages and such a grammaticalisation process makes us believe that it might have occurred as a drift in these three languages.

4.2 Parallel development and the grammaticalisation path of *ki*-V 'passive'

While *ki*-N is found in a number of Formosan languages, it is clear that it is most productive, i.e., *ki*- attaches to the widest range of nouns, in Rukai, Paiwan and Puyuma. We may hypothesise that while these three languages exhibit parallel constructions, *ki*-V 'passive' underwent similar pathways of grammaticalisation in these three languages, but independently and not through contact.

We show in the following section that the grammaticalisation path of *ki*-V 'passive' in Paiwan and Puyuma might differ from that in Rukai.

4.2.1 Grammaticalisation path of *ki*-V 'passive' in Paiwan and Puyuma

One reason that can be advanced for the parallel development of *ki*-V 'passive' in Paiwan and Puyuma is as follows. In these two languages, *ki-* can attach to nouns and to verbs, as well as to precategorial roots, which can be defined as

> lexical bases which do not occur without further affixation or outside a compound in any syntactic function and from which items belonging to different morphological or syntactic categories (nouns and verbs, for example) can be derived, without there being clear evidence that one of the possible derivations from a given root is more basic than the other one(s). (Himmelmann 2005:129).

(37) PUYUMA (Teng 2008)

a.	*tubang*	'answer, to answer'	*ki-tubang*	'get an answer, get answered'
b.	*sabung*	'compensation, to compensate'	*ki-sabung*	'get compensation, be compensated'
g.	*turu*	'warning, to warn'	*ki-turu*	'get a warning, be warned'

(38) NORTHERN PAIWAN

a.	*valra*	'authorise, permit'	*'i-valra*	'pay a call (i.e., get permission)'
b.	*tevelra*	'answer'	*'i-tevelra*	'get an answer'

We believe that in these two languages, the fact that *ki-* could attach to nouns and to precategorial roots may have led the way for the affixation of *ki-* with verb bases.

What appears more intriguing is the occurrence of a 'passive' in languages that have other well-established morphemes to indicate a speaker's viewpoint, namely AV versus NAV affixes. However, the function of *ki-* as a valency-changing morpheme does not overlap with the other voice markers in Paiwan and Puyuma.[15] Indeed, as we have shown above, *ki*-V and NAV-marked verbs contrast both syntactically and semantically.

On the syntactic level, we have demonstrated that:

(i) *ki*-Vs are syntactically intransitive, while NAV-marked verbs are syntactically transitive;

(ii) the undergoer (or patient argument) is marked as nominative in both *ki*-V and NAV clauses, but the marking of the agent argument differs: it is marked as oblique in Puyuma and in Paiwan in *ki*-V clauses; it is marked as oblique and cross-referenced on the verb with a genitive pronoun in Puyuma; and it is marked as genitive in Paiwan; and

(iii) the agent of a *ki*-V clause cannot function as a shared argument in a serial verb construction. It can in NAV clauses.

Semantically, the patient of *ki*-V clauses exhibit a higher degree of volition than that of NAV-clauses.

---

[15] Shibatani (1985:835) notes that '[i]t is not uncommon for a language to have two or more types of passives, but it is doubtful that they have the same distribution and function.'

4.2.2 Grammaticalisation path of *ki*-V 'passive' Rukai

While the development of *ki*-V 'passive' is only partial in Paiwan and in Puyuma, it has been fully grammaticalised in the Rukai dialects. For one thing, none of the Rukai dialects have reflexes of the voice markers which are found in all the other Formosan languages (PAn *<um>, *-en, *-an and *Si-/*Sa-). However, they do exhibit reflexes of these same forms as nominalising affixes (Zeitoun and Huang 2006). Thus, it can be hypothesised that at some point in the history of the Rukai dialects *ki-* took over functions usually carried out by NAV affixes. Based on recent findings by Malcolm Ross (pers.comm.), possible steps in the development of *ki*-V as a passive in Rukai are:

(i) *ki*-V evolved from *ki*-N as a passive;

(ii) its function expanded and it came to replace the PAn indicative verbal NAV affixes, cf. *-aw, *-ay and *-anay according to Ross' (2009) reconstruction;

(iii) consequently, Rukai never underwent — as most the other Formosan languages did (except Puyuma and Tsou) — the 'verb-from-nominalisation' shift, whereby the earlier PAn nominalising affixes *<um>, *-en, *-an and *Si-/ *Sa- were reanalysed as verbal affixes; rather in Rukai these affixes are preserved with their nominalising functions; and

(iv) another passive construction, the agentless passive, emerged in all the Rukai dialects except in Mantauran.

5 Conclusions

We have shown in this paper that the distribution of *ki*-N as 'obtain, get' in several Formosan languages supports its reconstruction for PAn. Its function as a passive in Rukai, Paiwan and Puyuma is not a product of a shared innovation, and thus we posit that the grammaticalisation of *ki*-N 'get N' as *ki*-V 'passive' occurred after the diversification of the Formosan languages. We postulated two hypotheses: (i) diffusion through contact, such that *ki*-V as 'passive' may have been borrowed from one language to another, explaining its similar function in geographically close languages; or (ii) parallel development, i.e., the grammaticalisation path of *ki*-V as 'passive' is a natural process that might have occurred in a parallel fashion in these three languages. We concluded that the latter hypothesis appears to be the most plausible, at this stage.

References

Blust, Robert. 1999a. 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 No.1., 31–94. Taipei: Academia Sinica.

—— 1999b. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2).321–365.

—— 2003. *Thao dictionary*. Language and Linguistics Monograph Series, A5. Taipei: Academia Sinica.

Cauquelin, Josiane. 1991. *Dictionnaire puyuma-français*. Textes et Documents Nousantariens IX. Paris: EFEO.

Chang, Hsiu-chuan. 2006. A grammar of Paiwan. PhD thesis. Canberra: The Australian National University.

Corne, Chris. 1977. *Seychelles Creole grammar: elements for Indian Ocean Proto-Creole reconstruction.* Tübingen: TBL-Verlag Narr.

Egerod, S. 1999. *Atayal–English dictionary*. (2nd edition edited by Jens Ostergaard Peterson.) Historisk-filosofiske Skrifter 20. Copenhagen: Det Kongelige Danske Videnskabernes Selskab [The Royal Danish Academy of Sciences and Letters].

Ferrell, Raleigh. 1969. *Taiwan Aboriginal groups: problems in cultural and linguistic classificatio*n. Institute of Ethnology Monograph No.17. Taipei: Academia Sinica.

—— 1982. *Paiwan dictionary*. Canberra: Pacific Linguistics.

—— 1983. Intent and volition in Paiwan and Tagalog verbs. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics,* vol. 4: *Thematic variation*, 1–8. Canberra: Pacific Linguistics.

Fey, Virginia. 1986. *Amis dictionary*. Taipei: The Bible Society.

Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.

Givón, Talmy and Lynne Yang. 1994. The rise of the English GET-passive. In Barbara Fox and Paul J. Hopper, eds *Voice: form and function*, 119–149. Amsterdam/Philadelphia: John Benjamins.

Heine, Bernd and Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

—— 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. In K. Alexander Adelaar and Nikolaus Himmelmann, eds *The Austronesian languages of Asia and Madagascar*, 110–181. London/New York: Routledge.

Ho, Dah-an. 1978. A preliminary comparative study of five Paiwan dialects. *Bulletin of the Institute of History and Philology* 49(4).565–681. [In Chinese]

——1983. The position of Rukai in the Formosan languages. *Bulletin of the Institute of History and Philology* 47(2).243–274. [In Chinese]

—— 1997. Kanakanavu. In Paul Jen-kuei Li, ed. *The Austronesian languages of Kaohsiung County*, 228-271. Kaohsiung County: Kaohsiung County government. [In Chinese]

Kaufman, Daniel. 2009. On the scope and function of PAn *<R> and *<ŋ>. Paper presented at the 11th International Conference on Austronesian Linguistics. Aussois, France.

Li, Paul Jen-kuei. 1973. *Rukai structure*. Institute of History and Philology, Special Publications No.64. Taipei: Academia Sinica.

—— 1997. Saaroa. In Paul Jen-kuei Li, ed. *The Austronesian languages of Kaohsiung County*, 272–297. Kaohsiung County: Kaohsiung County Government. [In Chinese]

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Language and Linguistics Monograph Series No. A2. Taipei: Academia Sinica.

—— 2006. *Kavalan dictionary*. Language and Linguistics Monograph Series A19. Taipei: Academia Sinica.

Lin, Tai, Si-qi Zeng, Wen-shu Li, and Bukun Ismahasan. 2001. *A study on the morphology of Isbukun Bunun*. Taipei: Duce Wenhua Publishing Company. [in Chinese].

Pecoraro, Ferdinando. 1979. *Essai de dictionnaire taroko-français*. Cahier d'Archipel 7. Paris: Société pour l'Etude de la Connaissance du Monde Insulindien (SECMI).

Ross, Malcolm. 1995. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng, eds *Austronesian studies relating to Taiwan*. Symposium Series of the Institute of History and Philology No. 3, 727–791. Taipei: Academia Sinica.

—— 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.

Ross, Malcolm and Stacy F. Teng. 2005. Formosan languages and linguistic typology. *Language and Linguistics* 6(4).739–781.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43.411–444.

Shibatini, Masayoshi. 1985. Passives and related constructions: a prototype analysis. *Language* 61(4).821–848.

Starosta, Stanley. 1994. Proto-Rukai-Tsouic: subgroup or treetop? Paper presented at the Seventh International Conference on Austronesian Linguistics. Leiden, The Netherlands.

—— 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, eds *Austronesian studies relating to Taiwan*, Symposium Series of the Institute of History and Philology No.3, 683–726. Taipei: Academia Sinica.

Starosta, Stanley, Andrew Pawley and Lawrence A. Reid. 1981. The evolution of focus in Austronesian. Unpublished manuscript. Honolulu: University of Hawai'i at Mānoa.

—— 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*, vol. 2: *tracking the travellers*, 145–170. Canberra: Pacific Linguistics.

Teng, Stacy F. 2008. *A grammar of Puyuma, an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Yeh, Marie Mei-li. 2003. A syntactic and semantic study of Saisiyat verbs. PhD thesis. Taipei: National Taiwan Normal University.

Van den Berg René. 2006. An unusual passive in Western Oceanic: the case of Vitu. Paper presented at the Tenth International Conference on Austronesian Linguistics. Palawan, The Philippines. [Available online: http://www.sil.org/asia/philippines/ical/papers.html]

Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. In Andrew B. Gonzales, ed. *Parangal Kay Cecilio Lopez (Essays in honor of Cecilio Lopez on his Seventy-fifth Birthday)*. Philippine Journal of Linguistics Special Monograph Issue No. 4, 71–94. Quezon City: Linguistic Society of the Philippines.

Zeitoun, Elizabeth. 1995. Problèmes de linguistique dans les langues aborigènes de Taiwan. [Issues on Formosan linguistics]. PhD thesis. Paris: Université René Diderot Paris 7.

—— 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Paper presented as keynote address at AFLA 10. Hawai‘i: University of Hawai‘i at Mānoa.

—— 2005. The internal relationships of Rukai revisited. Unpublished manuscript. Taipei: Academia Sinica.

—— 2007. *A grammar of Mantauran Rukai*. Language and Linguistics Monograph Series A4-2. Taipei: Institute of Linguistics, Academia Sinica.

—— in preparation. A comparative study of Rukai. Unpublished manuscript. Taipei: Academia Sinica.

Zeitoun, Elizabeth and Lillian M. Huang. 1997. Toward a typology of tense, aspect and modality in the Formosan languages: a preliminary study. In Chiu-yu Tseng, ed. *Chinese languages and linguistics*, vol. 4: *typological studies of languages in China*, 595–618. Taipei: Academia Sinica.

—— 2006. Nominalization in the Formosan languages. Paper presented at AFLA 13. Hsinchu: National Tsing Hua University.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual and modal systems of the Formosan languages: a typological perspective. *Oceanic Linguistics* 35(1).21–56.