A journey through Austronesian and Papuan linguistic and cultural space
Papers in honour of Andrew Pawley
Pacific Linguistics 615

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

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

FOUNDING EDITOR: Stephen A. Wurm
EDITORIAL BOARD: Malcolm Ross and I Wayan Arka (Managing Editors), Mark Donohue, Nicholas Evans, David Nash, Andrew Pawley, Paul Sidwell, Jane Simpson, and Darrell Tryon

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

edited by
John Bowden, Nikolaus P. Himmelmann and Malcolm Ross
with the editorial assistance of Edgar Suter

Pacific Linguistics
School of Culture, History and Language
College of Asia and the Pacific
The Australian National University
# Table of contents

<table>
<thead>
<tr>
<th>Chapter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
</tr>
<tr>
<td>Acknowledgements</td>
</tr>
<tr>
<td>List of Abbreviations</td>
</tr>
</tbody>
</table>

## ~ About Andy ~

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A special power</td>
</tr>
<tr>
<td>Byron W. BENDER</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Recollections of Andy Pawley</td>
</tr>
<tr>
<td>George W. GRACE</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>A collegial colleague to ANU linguists on ‘the other side’</td>
</tr>
<tr>
<td>Harold KOCH</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Half a century of our antipodean friendship</td>
</tr>
<tr>
<td>Viktor KRUPA</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>A brief (incomplete) history of the Oceanic lexicon project</td>
</tr>
<tr>
<td>Meredith OSMOND</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>Growing up with a failed anthropologist</td>
</tr>
<tr>
<td>Victor PAWLEY and Matthew PAWLEY</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>Andrew Pawley’s writings</td>
</tr>
<tr>
<td></td>
<td>21</td>
</tr>
</tbody>
</table>

## ~ Prologue ~

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>A metrical system that defies description by ordinary means</td>
</tr>
<tr>
<td>Alan RUMSEY</td>
<td>39</td>
</tr>
</tbody>
</table>

vii
~ Across the disciplines ~

9  Genes languages and agriculture in New Guinea population prehistory
   Robert ATTenBOROUGH  59

10 Language families and the history of human migration
    Peter BELLWOOD  79

11 Friends, relatives and enemies: the archaeology and history of
   interaction among Austronesian and non-Austronesian speakers
   in the western Solomons
   Peter SHEPPARD, Richard WALTER and Kenneth ROGA  95

12 ‘I was so much older then, I’m younger than that now’: why the dates
   keep changing for the spread of Austronesian languages
   Matthew SPRIGGS  113

13 The impact of a dynamic environmental past on trade routes and language
   distributions in the lower-middle Sepik
   Pamela SWADLING  141

~ Reconstructing language history ~

14 The amalgamation of Malagasy
   Alexander ADELAAR  161

15 Proto Oceanic *akin[i]: reconstructing change in progress
   Bethwyn EVANS  179

16 The extinct language of Vunaqumu
   Paul GERAGHTY  197

17 The outer eastern islands of the Solomons, a puzzle for the holistic
   approach to the anthropology of history
   Roger C. GREEN  207

18 Count not on substrata
   Peter C. LINCOLN  225

19 Lexical history in the Northwest Solomonic languages: evidence for two
   waves of Oceanic settlement in Bougainville and the northwest Solomons
   Malcolm ROSS  245

20 What can the Mon-Khmer lexical borrowings in Acehnese tell us?
   Paul SIDWELL  271
21 The languages of Espiritu Santo, Vanuatu
Darrell TRYON 283

22 A Tuamotuan challenge to the subgrouping of Hawaiian
William H. WILSON 291

~ Grammar and use ~

23 Numerals in Rongga
I Wayan ARKA 319

24 Bound roots in Balinese and Indonesian – precategorials or verbs?
Adrian CLYNES 333

25 Demonstratives in Tokelauan discourse
Robin HOOPER 357

26 Subject incorporation in Toqabaqita and other Oceanic languages
Frantisek LICHTENBERK 375

27 The fourth person in Teop
Ulrike MOSEL 391

28 Measures and morphemes: a functional approach to Hawaiian accent
Albert J. SCHÜTZ 405

29 The optional ergative in Kâte
Edgar SUTER 423

~ Complex predicates and complex sentences ~

30 Manner and path verb serialisation in Numbami
Joel BRADSHAW 441

31 Event integrity and continuity in the Tuam language
Robert D. BUGENHAGEN 455

32 Arie, Andy-ko! Aiyakoe bekâ resena!
Cynthia FARR 475

33 Wordhood in serial verb constructions: evidence from Anamuxra
Andrew INGRAM 481

34 Complex predicates and complex sentence types in Lavukaleve
Angela TERRILL 499
~ Lexicography, formulaic language and the lexicon ~

35  *Mauri, tupu* and the secret life of plant names
    Richard BENTON  515

36  Five patterns of semantic change in Austronesian languages
    Robert BLUST  525

37  Patterns of replacement in the Samoan lexicon
    ROSS CLARK  547

38  Language as fun and secret code: some play varieties in the Pacific
    Robert EARLY  555

39  Exploring oral formulaic language: a five poet analysis
    James J. FOX  573

40  New Zealand Māori plant names for teatree: manuka (*Leptospermum scoparium*) and kanuka (*Kunzea ericoides*), Myrtaceae
    Rhys GARDNER  589

41  Some thoughts on change in the Samoan language
    Cluny MACPHERSON  591

42  Proto Polynesian *kaina* 1. the populace of commoners;
    2. matrilineal descent group
    Jeff MARCK  607

43  Assaying the scope and number of fish names in Pendau based on Pawley’s Oceanic benchmarks
    Phil QUICK  621

44  Metaphors we judge by: mediation in Wehali
    Catharina WILLIAMS-VAN KLINKEN  633

45  Issues in the lexicographic presentation of Raga verb phrase grammatical functors
    D.S. WALSH  641

46  ‘Eating’ and ‘drinking’ in Kalam
    Anna WIERZBICKA  651

~ Epilogue ~

47  Taking up the ‘Pawley challenge’: speech formulae and linguistic theory
    Nikolaus P. HIMMELMANN  663
Preface

It is with great pleasure that we present this book to Andy Pawley, who has for each of us been a wise teacher, a respected colleague, and a very good friend. Conceiving a book like this one was a particularly difficult task. The first problem was that Andy — being such a gregarious and gracious fellow — has probably been the most prolific contributor to and editor of other people’s festscripts that we have known. No doubt he knows by name more people who have worked in the Coombs Building at The Australian National University than does anyone else with an office there, whether those people be scholars of international renown, students, administrators, or maintenance staff.

Asking all Andy’s friends and colleagues to contribute to a book in his honour would have given us a massive collection, probably running to several volumes. Clearly, that wasn’t going to work. We could have limited the contributors to only the most esteemed of his academic fellow-travellers, but that didn’t seem to be the best approach for Andy either, who has always been so welcoming and friendly to everyone he comes into contact with. In the end, we decided to restrict contributions to people who had been either colleagues or students within one of the institutions Andy had worked at: the University of Auckland, the University of Hawai’i, and The Australian National University.

But this restriction alone would also have created problems. We couldn’t ask all his former colleagues and students to contribute, because then the book would have had no focus at all, and would have had to cover all the disciplines studied in the Coombs Building, as well as a whole lot more. We couldn’t even just limit contributions to topics within Andy’s range of interests, as those interests are so wide-ranging. So it seemed best to limit the book to personal reminiscences of Andy and to the subjects the initial editors, John Bowden and Nikolaus P. Himmelmann, shared in common with Andy. The result is a journey through Austronesian and Papuan linguistic and cultural space. The journey will take the reader from song metrics to linguistic theory, along the way traversing the disciplines of genetics, archaeology, anthropology and, of course, linguistics, and visiting the linguistic subdisciplines of historical reconstruction, grammatical description and lexicography, all three of which are well represented in Andy’s bibliography.

The book has taken much longer to put together than we originally planned and hoped. Our original intention was to have the book completed by July 2007 for presentation at the Seventh Conference on Oceanic Linguistics in Nouméa, New Caledonia, so that as many as possible of Andy’s friends and colleagues could be present. And indeed, the presentation took place, but what was presented was a partial mock-up of the book, a ‘virtual’ version, so to speak. After this, things did not progress as fast as planned, and the third editor, Malcolm Ross, joined the team in the hope that the freedom he enjoyed in retirement would allow him to devote more time to the tasks that remained to be completed.
One consequence of the delayed publication is that three of the authors — Cindi Farr, Roger Green and Phil Quick — passed away before they could see their contributions in print. We can only say how sad we are that they are no longer with us.

Andy turned 65 in April 2006, and retired at the end of that year. The presentation took place at a time still close enough to be considered somehow related to these events. Publication comes considerably later so that the book now celebrates Andy’s work somewhere between 65 and 70, in the midst of his early retirement years.

We hope that Andy continues to enjoy a productive retirement for many years to come, and we hope that having escaped the admin he’ll find the time to complete what will be his crowning achievements, the dictionaries of Wayan Fijian and Kalam that he has been beavering away at for almost all his working life. We also hope that he finds this volume a fitting mark of so many people’s admiration.

Nō reira, tenā koe, tenā koe, tenā koe

John Bowden, Nikolaus P. Himmelmann and Malcolm Ross
March 2010
Acknowledgments

The original impetus for this festschrift for Andy Pawley came late in 2005 from a former student, Edgar Suter, also a contributor to this volume, who wrote to one of the editors asking whether anyone was editing a festschrift for Andy. Shortly after this, in February 2006, John Bowden wrote to former and current colleagues and students of Andy’s, inviting them to contribute to a festschrift. Edgar has continued to be an active participant in the book’s creation, providing a large measure of editorial assistance.

The editors would like to thank all the authors for their fine papers and reminiscences and for their patience during the tortuous process of production. We also thank the colleagues who served as anonymous referees. The bibliography of Andy’s works was edited from Andy’s curriculum vitae, and we are grateful to Meredith Osmond for this work.

We are indebted to Felicita Carr, Lila San Roque and Melissa Crowther for copy-editing the draft papers, to Jeanette Coombes for formatting the papers and overseeing the proofreading process, and to Julie Manley for administration and for the cover design.
**Abbreviations used in interlinear glosses**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>actor, most agentive argument of a transitive verb</td>
</tr>
<tr>
<td>ABL</td>
<td>ablative</td>
</tr>
<tr>
<td>ABS</td>
<td>absolutive</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative</td>
</tr>
<tr>
<td>ACDT</td>
<td>accidental</td>
</tr>
<tr>
<td>ACT</td>
<td>actor, active</td>
</tr>
<tr>
<td>ACTN</td>
<td>action particle</td>
</tr>
<tr>
<td>ADJ</td>
<td>adjective, adjective marker</td>
</tr>
<tr>
<td>ADV</td>
<td>adverb(ial)</td>
</tr>
<tr>
<td>ADVS</td>
<td>adversative passive</td>
</tr>
<tr>
<td>AFM</td>
<td>affirmation particle</td>
</tr>
<tr>
<td>AGT</td>
<td>agent</td>
</tr>
<tr>
<td>AL</td>
<td>alienable</td>
</tr>
<tr>
<td>ALL</td>
<td>allative</td>
</tr>
<tr>
<td>AN</td>
<td>animate</td>
</tr>
<tr>
<td>ANAPH</td>
<td>anaphoric</td>
</tr>
<tr>
<td>AND</td>
<td>andative</td>
</tr>
<tr>
<td>ANT</td>
<td>anterior</td>
</tr>
<tr>
<td>AOR</td>
<td>aorist</td>
</tr>
<tr>
<td>APP</td>
<td>applicative</td>
</tr>
<tr>
<td>AQ</td>
<td>indicative assertion</td>
</tr>
<tr>
<td>ART</td>
<td>article</td>
</tr>
<tr>
<td>AS</td>
<td>anticipatory subject</td>
</tr>
<tr>
<td>ASS</td>
<td>associated</td>
</tr>
<tr>
<td>ASSPL</td>
<td>associative plural</td>
</tr>
<tr>
<td>ATD</td>
<td>attitude/attitudinal</td>
</tr>
<tr>
<td>ATTR</td>
<td>attributive</td>
</tr>
<tr>
<td>AUX</td>
<td>auxiliary</td>
</tr>
<tr>
<td>AV</td>
<td>actor voice</td>
</tr>
<tr>
<td>BEN</td>
<td>benefactive</td>
</tr>
<tr>
<td>C</td>
<td>consonant</td>
</tr>
<tr>
<td>CAUS</td>
<td>causative</td>
</tr>
<tr>
<td>CEFF</td>
<td>marker on participant/event effecting change or in control (has agent/actor, force, and instrument roles)</td>
</tr>
<tr>
<td>CIRC</td>
<td>circumstantial</td>
</tr>
<tr>
<td>CLF</td>
<td>classifier</td>
</tr>
<tr>
<td>CMP</td>
<td>complementiser</td>
</tr>
<tr>
<td>CMPR</td>
<td>comparative</td>
</tr>
<tr>
<td>CNJ</td>
<td>conjunction</td>
</tr>
<tr>
<td>CNTP</td>
<td>contemplated</td>
</tr>
<tr>
<td>COIN</td>
<td>coincidental</td>
</tr>
<tr>
<td>COLL</td>
<td>collective</td>
</tr>
<tr>
<td>COM</td>
<td>comitative</td>
</tr>
<tr>
<td>CONC</td>
<td>concessive</td>
</tr>
<tr>
<td>COND</td>
<td>conditional</td>
</tr>
<tr>
<td>CONT</td>
<td>continuative</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>CPL</td>
<td>completive</td>
</tr>
<tr>
<td>CT</td>
<td>contrastive</td>
</tr>
<tr>
<td>CUST</td>
<td>customary</td>
</tr>
<tr>
<td>CV</td>
<td>conveyance voice</td>
</tr>
<tr>
<td>DADV</td>
<td>deictic adverbial</td>
</tr>
<tr>
<td>DPTCL</td>
<td>discourse particle</td>
</tr>
<tr>
<td>DAT</td>
<td>dative</td>
</tr>
<tr>
<td>DEF</td>
<td>definite</td>
</tr>
<tr>
<td>DEI</td>
<td>deictic</td>
</tr>
<tr>
<td>DEM</td>
<td>demonstrative</td>
</tr>
<tr>
<td>DEP</td>
<td>dependent</td>
</tr>
<tr>
<td>DES</td>
<td>desiderative</td>
</tr>
<tr>
<td>DEST</td>
<td>destinative</td>
</tr>
<tr>
<td>DET</td>
<td>determiner</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Meaning</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>DIM</td>
<td>diminutive</td>
</tr>
<tr>
<td>DIR</td>
<td>directional</td>
</tr>
<tr>
<td>DIST</td>
<td>distal</td>
</tr>
<tr>
<td>DISTR</td>
<td>distributive</td>
</tr>
<tr>
<td>DP</td>
<td>distant past</td>
</tr>
<tr>
<td>DO</td>
<td>direct</td>
</tr>
<tr>
<td>DS</td>
<td>different subject</td>
</tr>
<tr>
<td>DTR</td>
<td>detransitiviser</td>
</tr>
<tr>
<td>DU</td>
<td>dual</td>
</tr>
<tr>
<td>DUR</td>
<td>durative</td>
</tr>
<tr>
<td>DV</td>
<td>dative/directional voice</td>
</tr>
<tr>
<td>DYN</td>
<td>dynamic</td>
</tr>
<tr>
<td>ELAT</td>
<td>elative</td>
</tr>
<tr>
<td>EMP</td>
<td>empathy-seeking first person pronoun</td>
</tr>
<tr>
<td>EMPH</td>
<td>emphatic, emphasis</td>
</tr>
<tr>
<td>ERG</td>
<td>ergative</td>
</tr>
<tr>
<td>EVD</td>
<td>evidential</td>
</tr>
<tr>
<td>EVIT</td>
<td>evitative</td>
</tr>
<tr>
<td>EX</td>
<td>exclusive</td>
</tr>
<tr>
<td>EXCL</td>
<td>exclamation</td>
</tr>
<tr>
<td>EXIST</td>
<td>existential</td>
</tr>
<tr>
<td>EXP</td>
<td>experiential</td>
</tr>
<tr>
<td>EXT</td>
<td>extended</td>
</tr>
<tr>
<td>F</td>
<td>feminine</td>
</tr>
<tr>
<td>FAM</td>
<td>familiar</td>
</tr>
<tr>
<td>FN</td>
<td>future non-asserted</td>
</tr>
<tr>
<td>FOC</td>
<td>focus</td>
</tr>
<tr>
<td>FPST</td>
<td>far past</td>
</tr>
<tr>
<td>FRM</td>
<td>formal</td>
</tr>
<tr>
<td>FRQ</td>
<td>frequentative</td>
</tr>
<tr>
<td>FUT</td>
<td>future</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>GER</td>
<td>gerund</td>
</tr>
<tr>
<td>GV</td>
<td>goal voice</td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
</tr>
<tr>
<td>HAVE</td>
<td>stativising prefix</td>
</tr>
<tr>
<td>HES</td>
<td>hesitation marker</td>
</tr>
<tr>
<td>HON</td>
<td>honorific</td>
</tr>
<tr>
<td>HORT</td>
<td>hortative</td>
</tr>
<tr>
<td>HUM</td>
<td>human</td>
</tr>
<tr>
<td>IM</td>
<td>immediate (past/future event with immediate present relevance)</td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
</tr>
<tr>
<td>IN</td>
<td>inclusive</td>
</tr>
<tr>
<td>INAL</td>
<td>inalienable</td>
</tr>
<tr>
<td>INAN</td>
<td>inanimate</td>
</tr>
<tr>
<td>INCH</td>
<td>inchoative</td>
</tr>
<tr>
<td>INCPL</td>
<td>incompletive</td>
</tr>
<tr>
<td>IND</td>
<td>indicative</td>
</tr>
<tr>
<td>INDEF</td>
<td>indefinite</td>
</tr>
<tr>
<td>INESS</td>
<td>inessive</td>
</tr>
<tr>
<td>INF</td>
<td>infinitive</td>
</tr>
<tr>
<td>INGR</td>
<td>ingressive</td>
</tr>
<tr>
<td>INS</td>
<td>instrumental</td>
</tr>
<tr>
<td>INTJ</td>
<td>interjection</td>
</tr>
<tr>
<td>INTSF</td>
<td>intensifier</td>
</tr>
<tr>
<td>INVIS</td>
<td>invisible</td>
</tr>
<tr>
<td>INVOL</td>
<td>involuntary</td>
</tr>
<tr>
<td>IPF</td>
<td>imperfective</td>
</tr>
<tr>
<td>IRR</td>
<td>irrealis</td>
</tr>
<tr>
<td>ITER</td>
<td>iterative</td>
</tr>
<tr>
<td>IV</td>
<td>instrumental voice</td>
</tr>
<tr>
<td>JNT</td>
<td>joint action</td>
</tr>
<tr>
<td>K.o.</td>
<td>kind of</td>
</tr>
<tr>
<td>LIG</td>
<td>ligature</td>
</tr>
<tr>
<td>LIM</td>
<td>limitative, limiter</td>
</tr>
<tr>
<td>LK</td>
<td>linker, ligature</td>
</tr>
<tr>
<td>LN</td>
<td>locative noun</td>
</tr>
<tr>
<td>LOC</td>
<td>locative</td>
</tr>
<tr>
<td>LV</td>
<td>locative voice</td>
</tr>
<tr>
<td>LUD</td>
<td>ludling reversal</td>
</tr>
<tr>
<td>M</td>
<td>masculine</td>
</tr>
<tr>
<td>MDL</td>
<td>modal</td>
</tr>
<tr>
<td>MED</td>
<td>medial</td>
</tr>
<tr>
<td>MIN</td>
<td>minimal (number of pronoun)</td>
</tr>
<tr>
<td>MMT</td>
<td>movement</td>
</tr>
<tr>
<td>MOD</td>
<td>mood marker</td>
</tr>
<tr>
<td>MT</td>
<td>motion</td>
</tr>
<tr>
<td>MUT</td>
<td>mutated</td>
</tr>
<tr>
<td>N</td>
<td>neuter, noun, nominal</td>
</tr>
<tr>
<td>NAR</td>
<td>narrative (tense)</td>
</tr>
<tr>
<td>NDIST</td>
<td>near distal</td>
</tr>
<tr>
<td>NEAR</td>
<td>near tense</td>
</tr>
<tr>
<td>NEG</td>
<td>negation</td>
</tr>
<tr>
<td>NEWSIT</td>
<td>new situation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>NF</td>
<td>non-final (verb)</td>
</tr>
<tr>
<td>NFIN</td>
<td>non-finite</td>
</tr>
<tr>
<td>NFUT</td>
<td>non-future</td>
</tr>
<tr>
<td>NHORT</td>
<td>near hortative</td>
</tr>
<tr>
<td>NHUM</td>
<td>non-human</td>
</tr>
<tr>
<td>NMIN</td>
<td>non-minimal (number of pronoun)</td>
</tr>
<tr>
<td>NOLD</td>
<td>non-old information</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative</td>
</tr>
<tr>
<td>NOW</td>
<td>now (deictic)</td>
</tr>
<tr>
<td>NP</td>
<td>noun phrase</td>
</tr>
<tr>
<td>NPST</td>
<td>near past</td>
</tr>
<tr>
<td>NR</td>
<td>nominaliser</td>
</tr>
<tr>
<td>NREF</td>
<td>non-referential</td>
</tr>
<tr>
<td>NSBJ</td>
<td>non-subject</td>
</tr>
<tr>
<td>NSG</td>
<td>non-singular</td>
</tr>
<tr>
<td>NSPEC</td>
<td>non-specific</td>
</tr>
<tr>
<td>NTRL</td>
<td>neutral distance (demonstrative category)</td>
</tr>
<tr>
<td>O</td>
<td>most patient-like argument of a transitive verb</td>
</tr>
<tr>
<td>OBJ</td>
<td>object</td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
</tr>
<tr>
<td>OLD</td>
<td>old information</td>
</tr>
<tr>
<td>OPT</td>
<td>optative</td>
</tr>
<tr>
<td>ORD</td>
<td>ordinal number</td>
</tr>
<tr>
<td>OV</td>
<td>objective voice</td>
</tr>
<tr>
<td>P</td>
<td>possessive marker</td>
</tr>
<tr>
<td>PART</td>
<td>participle</td>
</tr>
<tr>
<td>PASS</td>
<td>passive</td>
</tr>
<tr>
<td>PAT</td>
<td>patient</td>
</tr>
<tr>
<td>PAU</td>
<td>paucal</td>
</tr>
<tr>
<td>PERL</td>
<td>perlative</td>
</tr>
<tr>
<td>PERS</td>
<td>personal</td>
</tr>
<tr>
<td>PERSMKR</td>
<td>personal marker</td>
</tr>
<tr>
<td>PET</td>
<td>petitive</td>
</tr>
<tr>
<td>PF</td>
<td>perfect, perfective</td>
</tr>
<tr>
<td>PL</td>
<td>plural</td>
</tr>
<tr>
<td>PM</td>
<td>predicate marker</td>
</tr>
<tr>
<td>PN</td>
<td>proper name</td>
</tr>
<tr>
<td>POSS</td>
<td>possessive</td>
</tr>
<tr>
<td>PO</td>
<td>postverbal particle</td>
</tr>
<tr>
<td>POT</td>
<td>potentive (potential)</td>
</tr>
<tr>
<td>PP</td>
<td>prepositional phrase</td>
</tr>
<tr>
<td>PPR</td>
<td>present progressive</td>
</tr>
<tr>
<td>PR</td>
<td>presentative</td>
</tr>
<tr>
<td>PRE</td>
<td>preverbal particle</td>
</tr>
<tr>
<td>PREC</td>
<td>precedentive</td>
</tr>
<tr>
<td>PRD</td>
<td>predicative (pronoun, deictic)</td>
</tr>
<tr>
<td>PRED</td>
<td>predicate</td>
</tr>
<tr>
<td>PREP</td>
<td>preposition</td>
</tr>
<tr>
<td>PRG</td>
<td>progressive</td>
</tr>
<tr>
<td>PRH</td>
<td>prohibitive</td>
</tr>
<tr>
<td>PROFORE</td>
<td>pronominal foregrounder</td>
</tr>
<tr>
<td>PRON</td>
<td>pronoun</td>
</tr>
<tr>
<td>PRPV</td>
<td>propositive</td>
</tr>
<tr>
<td>PRTV</td>
<td>partitive</td>
</tr>
<tr>
<td>PRX</td>
<td>proximal</td>
</tr>
<tr>
<td>PST</td>
<td>past tense</td>
</tr>
<tr>
<td>PTCL</td>
<td>particle</td>
</tr>
<tr>
<td>PURP</td>
<td>purposive</td>
</tr>
<tr>
<td>PV</td>
<td>patient voice</td>
</tr>
<tr>
<td>Q</td>
<td>question (particle)</td>
</tr>
<tr>
<td>QUOT</td>
<td>quotative</td>
</tr>
<tr>
<td>R</td>
<td>realis</td>
</tr>
<tr>
<td>RCP</td>
<td>reciprocal</td>
</tr>
<tr>
<td>RCT</td>
<td>recent (deictic)</td>
</tr>
<tr>
<td>RDP</td>
<td>reduplication</td>
</tr>
<tr>
<td>RECOG</td>
<td>recognitionally deictic</td>
</tr>
<tr>
<td>RECX</td>
<td>recipient–experiencer</td>
</tr>
<tr>
<td>REL</td>
<td>relative (particle/affix), relative clause marker</td>
</tr>
<tr>
<td>REM</td>
<td>remote (deictic)</td>
</tr>
<tr>
<td>RES</td>
<td>resultative</td>
</tr>
<tr>
<td>RESID</td>
<td>residual</td>
</tr>
<tr>
<td>RFL</td>
<td>reflexive</td>
</tr>
<tr>
<td>RPST</td>
<td>remote past</td>
</tr>
<tr>
<td>RPRT</td>
<td>reportative</td>
</tr>
<tr>
<td>RQV</td>
<td>requestive</td>
</tr>
<tr>
<td>RSN</td>
<td>reason, (inanimate) cause</td>
</tr>
<tr>
<td>S</td>
<td>single argument of an intransitive verb</td>
</tr>
<tr>
<td>SBD</td>
<td>subordinate</td>
</tr>
<tr>
<td>SBJ</td>
<td>subject (marker)</td>
</tr>
<tr>
<td>SEQ</td>
<td>sequential, sequencing/ anterior medial verb</td>
</tr>
<tr>
<td>SER</td>
<td>serializing conjunction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SF</td>
<td>stem-forming morph</td>
</tr>
<tr>
<td>SG</td>
<td>singular</td>
</tr>
<tr>
<td>SIM</td>
<td>simulative, simultaneous</td>
</tr>
<tr>
<td>SOF</td>
<td>softening particle</td>
</tr>
<tr>
<td>SPEC</td>
<td>specific (article)</td>
</tr>
<tr>
<td>SQ</td>
<td>sequential conjunction</td>
</tr>
<tr>
<td>SR</td>
<td>subordinator</td>
</tr>
<tr>
<td>SS</td>
<td>same subject</td>
</tr>
<tr>
<td>ST</td>
<td>stative</td>
</tr>
<tr>
<td>SUBJ</td>
<td>subjunctive</td>
</tr>
<tr>
<td>SUCC</td>
<td>successive</td>
</tr>
<tr>
<td>SUP</td>
<td>superlative</td>
</tr>
<tr>
<td>SV</td>
<td>stem vowel</td>
</tr>
<tr>
<td>SVC</td>
<td>serial verb construction</td>
</tr>
<tr>
<td>TA</td>
<td>tense/aspect marker/particle</td>
</tr>
<tr>
<td>TAM</td>
<td>tense/aspect/mood marker</td>
</tr>
<tr>
<td>TEL</td>
<td>telic</td>
</tr>
<tr>
<td>THC</td>
<td>thematic consonant</td>
</tr>
<tr>
<td>TOP</td>
<td>topic</td>
</tr>
<tr>
<td>TP</td>
<td>today's past tense</td>
</tr>
<tr>
<td>TR</td>
<td>transitive, transitiviser</td>
</tr>
<tr>
<td>TRL</td>
<td>trial</td>
</tr>
<tr>
<td>TRUNC</td>
<td>truncation</td>
</tr>
<tr>
<td>UG</td>
<td>undergoer</td>
</tr>
<tr>
<td>UV</td>
<td>undergoer voice</td>
</tr>
<tr>
<td>V</td>
<td>vowel, verb</td>
</tr>
<tr>
<td>VC</td>
<td>verb complex</td>
</tr>
<tr>
<td>VENT</td>
<td>ventive</td>
</tr>
<tr>
<td>VIS</td>
<td>visible</td>
</tr>
<tr>
<td>VN</td>
<td>verbal noun</td>
</tr>
<tr>
<td>VOL</td>
<td>voluntary</td>
</tr>
<tr>
<td>VSF</td>
<td>verb stem forming formative</td>
</tr>
<tr>
<td>X, Y</td>
<td>used with arguments to denote a semantic role</td>
</tr>
<tr>
<td>1</td>
<td>1st person</td>
</tr>
<tr>
<td>2</td>
<td>2nd person</td>
</tr>
<tr>
<td>3</td>
<td>3rd person</td>
</tr>
<tr>
<td>1s</td>
<td>1st person singular</td>
</tr>
<tr>
<td>2s</td>
<td>2nd person singular</td>
</tr>
<tr>
<td>3s</td>
<td>3rd person singular</td>
</tr>
<tr>
<td>3sf</td>
<td>3rd person singular feminine</td>
</tr>
<tr>
<td>3sm</td>
<td>3rd person singular masculine</td>
</tr>
<tr>
<td>3sn</td>
<td>3rd person singular neuter</td>
</tr>
<tr>
<td>1d</td>
<td>1st person dual</td>
</tr>
<tr>
<td>1di</td>
<td>1st person dual inclusive</td>
</tr>
<tr>
<td>1de</td>
<td>1st person dual exclusive</td>
</tr>
<tr>
<td>2d</td>
<td>2nd person dual</td>
</tr>
<tr>
<td>3d</td>
<td>3rd person dual</td>
</tr>
<tr>
<td>1t</td>
<td>1st person trial</td>
</tr>
<tr>
<td>1ti</td>
<td>1st person trial inclusive</td>
</tr>
<tr>
<td>1te</td>
<td>1st person trial exclusive</td>
</tr>
<tr>
<td>2t</td>
<td>2nd person trial</td>
</tr>
<tr>
<td>3t</td>
<td>3rd person trial</td>
</tr>
<tr>
<td>1p</td>
<td>1st person plural</td>
</tr>
<tr>
<td>1pi</td>
<td>1st person plural inclusive</td>
</tr>
<tr>
<td>1pe</td>
<td>1st person plural exclusive</td>
</tr>
<tr>
<td>2p</td>
<td>2nd person plural</td>
</tr>
<tr>
<td>3p</td>
<td>3rd person plural</td>
</tr>
<tr>
<td>3s/p</td>
<td>subject pronoun referring to a singular or plural participant</td>
</tr>
<tr>
<td>4p</td>
<td>fourth person plural object pronoun</td>
</tr>
<tr>
<td>4s</td>
<td>fourth person singular object pronoun</td>
</tr>
<tr>
<td>4s/p</td>
<td>fourth person object pronoun referring to a singular participant in a clause with a third person plural subject</td>
</tr>
<tr>
<td>I</td>
<td>verb stem I</td>
</tr>
<tr>
<td>II</td>
<td>verb stem II</td>
</tr>
<tr>
<td>-</td>
<td>morpheme break</td>
</tr>
<tr>
<td>=</td>
<td>juncture of word and clitic</td>
</tr>
<tr>
<td>+</td>
<td>word juncture</td>
</tr>
<tr>
<td>?</td>
<td>meaning unknown</td>
</tr>
</tbody>
</table>
~ About Andy ~
A special power

BYRON W. BENDER

I am happy to be able to join in this celebration of the career of Andrew K. Pawley. Other contributions to this volume give us a glimpse of the content of the intellectual currents that he has helped shape — a bit of their breadth, depth, and flavour. I want to mention something that might more appropriately be put into a time capsule until the principals are no longer around to recognise themselves, a personal quality of Andrew’s¹ that I will refer to as his nose for talent. I do not claim sole proprietorship of this insight; I believe that those who chose him to head the department at RSPAS must have spotted it, too.

There is something uncanny about this quality, this special power of Andrew’s. It includes not only the ability to recognise unrealised talent, but the interpersonal skills — including a great sense of humour — that enable him to do something about it, to so influence events, and those in charge, that the situation is modified in a positive way. I will mention just three instances that I am aware of, although I know that they could be multiplied many times over. Each involves either the reclamation of a scholar, or the redeployment of one into a position better suited for realising his or her full potential. I will try to use some discretion, and possibly some poetic license, to provide minimal disguises.

• Scholar A, although already the author of two seminal papers, had never received the doctorate, and through a series of personal reverses found him/herself stuck in a humdrum blue collar job. The next thing we knew s/he had been granted a graduate assistantship at ANU whereby s/he was able to re-enter a life of productive scholarship.

• Scholar B applied for a position at the same institution that had granted her/him the doctorate some years previously. Since then s/he had held post docs at several other institutions that were leaders in the field, and seemed destined for a brilliant career. But his/her alma mater had qualms about hiring their own graduate. Then Andrew came on the scene and convinced everyone that this was indeed a valid exception. The rest is history, as the cliché has it, and the history is still being written.

¹ I refer to him throughout this tribute as Andrew, even though I generally address him as Andy. I continue to hear in my mental ear Medina’s referring to him and addressing him as Andrew. I don’t think I have ever heard her do otherwise. I hope she won’t mind my borrowing, for this special purpose, this device so laden with affection.
Scholar C was author of a slightly derivative, but nevertheless prominently and positively known book that helped interpret a new paradigm that was sweeping the discipline. S/he was in many ways a mature scholar, and although s/he had completed his/her own doctorate fairly recently, the new notoriety enabled him/her to be hired into a university department at a higher rank than was usually accorded recent PhDs. This created a certain backlash among his/her colleagues who were later called upon to decide upon his/her tenure. They equivocated and the decision was postponed for a year. Enter Andrew upon the scene. Not only was tenure granted, but Scholar C eventually became — after a productive faculty career — the college dean who helped preserve the department’s faculty lines through a series of retirements, so that the department was able to hire a new cohort of promising scholars.

Finally, those of us who have known Andrew as friend also know that he is a true friend. A mutual friend, Don Topping, who succumbed to cancer in 2003 after a 15-year bout, knew Andrew as someone who was always there for him, and who kept in touch through trans-Pacific phone calls up until the very end. I know that Don joins me in spirit in this tribute to Andrew.
It’s impossible to find a simple description for my relationship with Andy; it’s been both long and many-faceted. I first met him in Honolulu at or about the beginning of 1964 when Bruce Biggs and I were colleagues in an East-West Center project in which Andy was participating as Bruce’s student. A few years later, we were colleagues in the University of Hawai’i faculty for a time. And we have had a couple of formal collaborations. We were co-principal investigators of the NSF-funded New Guinea Oceanic project in about 1975–76, we co-taught a course in the 1977 LSA Institute in Hawai’i, and we have from time to time discussed other possible joint undertakings that didn’t materialise.

Although Andy was Bruce’s student at the time of our first encounter in 1964, I also played a significant part in shaping his future. As evidence I offer the fact that I had a young assistant named Medina Asuncion working for me at the time. Can I take credit for introducing her to Andy? It was a long time ago and I can’t remember for sure, but it really doesn’t matter — what we can be sure of is that Andy contrived in some manner to make her acquaintance. (And here the much-worn cliché ‘the rest is history’ suggests itself.)

Actually, as best I now remember, their courtship was conducted pretty much below the radar. When they got married, I was surprised — I hadn’t been sure they were actually aware of each other. (Well, that’s how I now remember it.)

Anyway with such a beginning, it may not be too surprising that I’ve followed Andy’s subsequent career with considerable personal interest even though he was never actually one of my students. But then I would no doubt have remained very aware of what he was doing even without this initial stimulus because our linguistic interests have followed such parallel lines.

In fact, although I don’t remember that I ever specifically discussed these matters with Andy, I think we’ve always had very similar overall views of what we’re trying to do that help put the more specific parallels into context. I believe, for example, that we have both seen ourselves as anthropologists — not in addition to being linguists, but as part of the same whole. We’ve both thought of linguistic genetic classification and reconstruction as part of culture-historical reconstruction. And beyond that, we’ve seen a more general need to integrate linguistic evidence together with other evidence to attack broader (not exclusively linguistic) problems.
The more specific parallel interests that I have in mind can be grouped without too much distortion under two heads: (1) historical reconstruction via language and (2) what I might describe as how language is really used and what that says about what speakers need to know.

Historical reconstruction via language, especially as applied to Oceanic languages, was my foremost interest at the time we met. Andy was certainly also involved with Oceanic languages although I no longer remember to what extent there was already a historical-comparative component to his interest. In any case, there soon was one.

As I look back on it, a fair assessment might be that Andy has tended to move on to further advances while I remained bogged down in details. While I was absorbed in trying to figure out what was wrong with our theories of linguistic change and differentiation, Andy was successfully applying them and producing results. He was taking a leading role both in narrowly linguistic reconstruction and in culture-historical reconstruction in collaboration with other disciplines. More specifically, he led the way in grammatical reconstruction — at least as far as the Oceanic branch of the family is concerned — and then, in collaboration with Roger Green, Malcolm Ross, and others, he has had considerable success in combining archeological, ethnological, and linguistic evidence in the reconstruction of culture history.

His more recent comparative work on the Trans New Guinea Phylum might also be mentioned under this head: historical reconstruction via language extended in this case to non-Austronesian languages.

If I remember correctly it was some years later, while he was at the University of Hawaii, that we became aware of another developing interest that we shared — in what I’m describing as how people actually use language. This, of course, has implications about what it is to know a language, which in turn offers new perspectives on all sorts of things including how languages differ, change, are stored in the brain, interact, can be translated, and much more.

Although we came to this new area of interest from quite different directions and were motivated by quite different concerns, I’ve learned a great deal from subsequent discussions with Andy, as well as from much that he’s written on the subject (or subjects, if one prefers). My interest developed in large part out of concerns about what I’ve referred to as ‘aberrant’ Austronesian languages — languages whose development from the Austronesian proto-language is difficult to reconstruct except for isolated details (and, to be fair, I suppose they all qualify as aberrant in some degree). Was it perhaps not true that languages consistently change in the regular ways we assumed? Or were there (as had been proposed, especially for Melanesia) ‘mixed’ languages that could not be derived from a single proto language at all? In fact, what actually was ‘language contact’?

A field survey that I’d carried out in many parts of Melanesia in 1955–56 had already made me think that our linguistic descriptions didn’t give us the right information for understanding the way contact actually affected languages. Subsequently I found that even a superficial examination of Sprachbünde such as the classical Balkans case confirmed that traditional linguistic descriptions provide no basis for explaining the convergences that actually occur. That, and work by John Gumperz and Louis Hjelmslev in particular, led me to the conclusion that so-called grammatical convergence was a by-product of what could be somewhat more revealingly described as a convergence in the way things were said (i.e. were construed for saying). What this convergence was leading toward specifically was morpheme-to-morpheme intertranslatability, and the key was to be found in what I called
‘idiomatology’. In particular, each language has its own way of saying a thing and generally of talking about the same subject, and this is part of what knowing the language consists of. Thus, there are implications reaching far beyond language convergence.

Now imagine my surprise to learn that Andy, working on English language data and with entirely different objectives, had arrived at very parallel conclusions. As would be expected, Andy’s route was much more hands-on. His interest was at least in part stimulated by his mother and collaborator, Frances Syder. She was concerned about the difference between the language skills that were taught in the schools and those required for natural conversation, a topic that Andy has also written on. Anyway, they set out to provide a better description of the rules of natural conversation in English, and in order to do so, collected and analysed a sizable corpus of conversations occurring in natural, informal settings. From my perspective, their findings were not only an encouraging confirmation of the conclusions I was reaching but also a major source of new perspectives and ideas.

I found what they’ve written in explanation of fluency and native-like selection — neither of which had even been recognised as a problem by the linguistic mainstream — a welcome addition to my developing understanding. The same may be said of their ‘one clause at a time hypothesis’. And of course Andy has contributed extensively to our understanding of formulaic speech. But one idea that I’m particularly eager to see developed further concerns what Andy has called ‘subject matter codes’. The idea is that our normal discourse is informed, indeed governed, by a shared understanding of how the particular subject matter is to be talked about (which also means a shared understanding of how it’s to be thought about). Andy had, and I hope still has, plans for a book on the subject.

The topic of subject matter codes brings me to Kalam. It seems clear that some languages are more ‘exotic’ than others from the perspectives of European languages (and of most linguists and of the assumptions of linguistic theory). Americanists have long sought to call the attention of other linguists to this exotic character in many Native American languages. In fact, it seems that from the perspective of any particular language there will be certain languages that seem much more exotic in this sense than others. However, the standard models of linguistic description have, as far as I know, provided no coherent picture of what this relative exoticism consists of.

Kalam, on which Andy has worked extensively, is like many other Papuan languages also quite exotic (in the same ethnocentric sense). And for this reason it has provided a particularly fertile ground for the analysis of the differences between ways of talking, and specifically of subject matter codes, in different languages. Andy has written several papers on this general subject — papers that advance our knowledge of subject matter codes while at the same time providing insight about what (relative) exoticism means.

It has to be quite some years now since Andy first began talking about writing the subject matter code book. I encouraged him from the beginning, and for some years every time I saw him, we went through the same routine: I asked about the progress of the book and he replied that he’d had no time (which in Andy’s case was always believable).

In the last decade or so I’ve had only occasional brief conversations with Andy when he’s passed through Honolulu. Fortunately, since he’s still pretty good at sending me offprints of his work, I feel I’ve kept fairly well abreast of what he’s been doing. This is not too difficult because, even though he of course continues to produce a great deal, what
he writes is really very readable. (I remember his once citing Stephen Jay Gould as a model for authors, but have never tried to compare their styles).

I’m hoping that his retirement will give him time finally to turn his attention to the work on subject matter codes. In fact, it’s especially for this reason that I feel grateful for the wisdom of the Australian government that allows him the freedom of retirement now when so much of his productive career lies ahead.

Before ending, I should perhaps make at least passing mention of the fact that outside of more scholarly pursuits we’ve shared an interest in sports, although I must admit I have nothing approaching Andy’s memory for actual events and record performances and know next to nothing about cricket. (Should that be an embarrassing admission for one who has actually once or twice been addressed as ‘Dr W.G. Grace’?) As far as I can remember our only actual sports encounter was a set or so of doubles, probably at the Diamond Head Tennis Center in Honolulu (where I seem to remember that Andy’s partner was Doug Yen; I can’t recall mine).

Finally, as I read what I’ve written here, there are a number of places where I wonder if Andy’s recollection will completely match my own. It’s true that my memory sometimes seems a bit unreliable — I forget a lot of things that happened and (I am told) remember some that didn’t. However, I’m sure that Andy will recognise the well-known and use-honored principle, ‘[history] is what you can remember’, and appreciate its applicability here.
A collegial colleague to ANU linguists on ‘the other side’

HAROLD KOCH

In these reminiscences I would like to reflect on Andrew Pawley’s involvement with linguistic colleagues on the ‘other side’ of The Australian National University; that is those of us in the Faculty of Arts.

To appreciate this perspective it is important for readers to understand something of the organisational structure of the ANU during the period of Andrew’s career in this institution (it is undergoing considerable restructuring at the present time). The university had a bipartite structure consisting of the Institute of Advanced Studies (IAS) with its research schools on the one hand and The Faculties on the other. Academic staff in the IAS were responsible primarily for research and the supervision of higher degree research (i.e. PhD and MPhil) students. Staff in The Faculties were engaged in teaching of undergraduate and post-graduate coursework students, but were involved in research and research student supervision as well. Andrew headed a Department of Linguistics in the Research School of Pacific and Asian Studies within the IAS. There was another Linguistics Department in the Faculty of Arts within The Faculties, which consisted of seven to ten linguists. This department was headed by its foundation professor R.M.W. (Bob) Dixon from 1970 until 1990, when he stepped down to take up a professorial research fellowship within the Faculty of Arts but funded by the Australian Research Council until his departure for La Trobe University in 2000. The Linguistics Department was subsequently headed by more junior staff members (myself, Tim Shopen, and Anthony Liddicoat) from 1991 until it was merged into a School of Language Studies (SLS), along with Modern European Languages and Classics, from 2001.

I am happy to report that since Andrew’s appointment to the Linguistics Department of RSPAS the relations between the linguists in RSPAS and those in the Faculty of Arts have always been cordial. During most of this period there was a Graduate School, which had oversight of ‘graduate programs’ that were organised along disciplinary lines. Each graduate program had a convenor who was appointed by a dean of graduate studies and who was responsible, with the assistance of a Board of Studies (consisting of the other staff

1 The principal reorganisation of the last few years has been the creation of colleges which span the IAS-Faculties divide: RSPAS belonged to the College of the Asian Pacific (CAP), along with the Faculty of Asian Studies; the Faculty of Arts was grouped with the Research School of Social Sciences (and a new Research School of Humanities) into a College of Arts and Social Sciences (CASS).

2 This was facilitated no doubt by the interest in cricket that Bob Dixon shared with Andrew.
Harold Koch

in the program), for managing the students in the relevant graduate program. Although the graduate program extended across both sides of the university, graduate students were located administratively within an appropriate department of either the IAS or The Faculties. The convenorship of the Linguistics Graduate Program alternated between the two Linguistics departments.

Andrew took his turn in serving as the Graduate Program Convenor over a two-year period (1999–2000). Former graduate students remember with fondness Andrew’s kindness, humility, approachability, and generosity with his time. Each PhD student needed a supervisory committee of three academics. Andrew’s broad range of expertise and his co-operative attitude resulted in his involvement in co-supervision of several students from our department — including Evershed Amuzu, Radu Daniliuc, and Marie Fellbaum.

A number of PhD students moved to Andrew’s department to pursue their PhD after doing coursework study in our department. These were students whose thesis topics were on languages for which Andrew and his colleagues had areal expertise that was lacking in our department. These topics included descriptions by Dorothy Jauncey and Catriona Hyslop of languages of Vanuatu, Deborah Hill and Angela Terrill of languages of the Solomon Islands, Louise Baird and Francisca Hondoko of languages of Indonesia, Tatsuya Yanagida and Carol Priestley of languages of Papua New Guinea; Bethwyn Evans on comparative Oceanic linguistics; and Bevan Barrett and Jacinta Smallhorn on topics of comparative Papuan linguistics. Andrew has sometimes joked that a function of our department was to supply graduate students to his department.

I would like to emphasise especially Andrew’s contribution to the teaching of (undergraduate) courses — since this was relatively rare for staff of the IAS. Over the period 1993–2006 Andrew consistently taught, about every two years, a course called Dictionaries and Dictionary-Making (LING2024). This was modelled on a course he had earlier taught at the University of Auckland. In 2005 he and his colleague Malcolm Ross taught an original course on Papuan comparative linguistics under the rubric of our course Study of a Language Family (LING3008). Their Papuan course was one of the best-attended of our annual courses on language families. As well as teaching courses of which he was the co-ordinator, Andrew has willingly given guest lectures in other courses: for courses in Austronesian linguistics co-ordinated by either our staff or (later) by his RSPAS colleagues, for Anna Wierzbicka’s Cross-Cultural Communication, for Johanna Rendle-Short’s Language and Society, and for my Language Change and Linguistic Reconstruction. In the latter I have valued his demonstration of the methods of reconstructing prehistoric culture from the evidence of language, using data from the project *The lexicon of Proto Oceanic*.

In 1995 ANU linguists were asked by the Australian Linguistic Society at relatively short notice whether we could organise the Australian Linguistic Institute (a batch of intensive courses over two weeks) in July of 1996. Andrew consented to act as the director of this ALI, leading the organising committee which put on an interesting suite of courses.

Another kind of collaboration across campus concerns the Australian National Dictionary Centre (ANDC). Andrew served as an ex officio member (as head of a linguistics department) of its Advisory Committee for a number of years. In fact, soon after his arrival at ANU Andrew chaired a committee reviewing the ANDC. This committee recommended, inter alia, that the heads of the two linguistics departments should be
members of the ANDC’s Advisory Committee and that a course in lexicography should be established (which he later consented to teach — as mentioned above).

One of the cross-campus collaborative endeavours of linguists at ANU concerns the Centre for Research on Language Change (see website http://crlc.anu.edu.au). This arose out of a seminar series in historical linguistics that I began in 1999. Andrew was one of the proponents of organising ourselves into a (virtual) centre — which came to fruition in 2001. One of the activities of the CRLC is a publication series called Studies in Language Change, whose existence was facilitated by the fact that it is a sub-series of the Pacific Linguistics series of which Andrew was at the time Chair of the Editorial Board.

While RSPAS linguists have always focused on specific languages to a greater degree than colleagues in Linguistics (Arts), Andrew has nevertheless contributed to research projects of colleagues in Arts. In particular, he contributed to a 1994 book on semantic and lexical universals edited by Cliff Goddard and Anna Wierzbicka a chapter on semantic primitives in Kalam, a Papuan language which he has made famous. Andrew has also lent advice and support to an ARC-funded project ‘Tracing change in family and social organisation in Indigenous Australia, using evidence from language’, on the reconstruction of kinship terminology and systems, which is currently being conducted in the School of Language Studies by myself, Patrick McConvell, anthropologist Ian Keen, plus a number of overseas collaborators (see website http://austkin.pacific-credo.fr/).

Andrew’s helpfulness and encouragement to other researchers are well-known. He has cheerfully provided useful feedback to a number of us linguists in the Faculty of Arts on grant proposals. In summary, Andrew has provided to us his colleagues a rare model of unselfish and generous scholarly collegiality, for which we thank him heartily.

Exotic reveries of life utterly different from earthbound unexciting reality often engage our attention at a certain age, especially when we feel cut off from the rest of the world and when, in addition to this, the borders of ‘reality’ coincide with the seemingly impenetrable and eternal Iron Curtain. Dreaming is a virtual escape in such circumstances when you are not willing to swim across a Danube guarded day and night by armed soldiers ready to shoot. This was how it was in Slovakia before 1989. For me, exotic dreams and yearnings were stimulated by books and movies, but mainly by books. In the late fifties of the 20th century I managed to find a selection of solid publications on Oceania in the libraries of the Philosophical Faculty in Bratislava — for example *Myths and songs from the South Pacific* by W.W. Gill, *Samoa a hundred years ago and long before* by George Turner or *Maori Life in Ao-Tea* by Johannes C. Andersen and the like. But no tide of new books from abroad. The borders were guarded like those of Japan during the two and half centuries of *sakoku*.

When I was still a student of general linguistics and Slavic studies, I decided to write a letter to the University of Auckland with the intention of finding a pen-friend in New Zealand who would be willing to correspond with me about Māori language and culture. Andy Pawley, several years younger than myself, replied. At that time he was studying at Auckland University. His reaction was generous — it was the reaction of the extremely kind person that he is to this day. He sent me — without delay — my first textbook of the Māori language, written by W.H. Wills, teacher of Māori at Te Aute College. The author of the book was not mistaken when he claimed that the students would advance very rapidly and easily using his handbook, published just two years earlier, in 1956: he was right, as a beginner, I can confirm it. Later *First lessons in Maori* (11th edition) by W.L. Williams and H.W. Williams followed. Contacts between Czechoslovakia and Western Europe started to improve, I acquired literature on Samoa, the Marquesas, Tonga, and in March 1959 Andy Pawley sent me the *Dictionary of the Maori language* by Herbert W. Williams (1957 edn.).

Our correspondence continued and I could keep an eye on what was going on at Andy’s course of Māori. In time I learned from him about his fieldwork among the Kalam in the highlands of Papua New Guinea, and about his research in Fiji and elsewhere. His ability to concentrate and yet to cover a broad linguistic expanse as he advanced from descriptive linguistics to the problems of historical linguistics, was and remains worthy of admiration.
In the meantime Andy married Medina. I had been granted a scholarship by the Wenner–Gren Foundation in New York and was looking forward to meeting Andy in Auckland. I stayed in New Zealand from December 1965 till March 1966 and could avail myself of his generous hospitality. He introduced me not only to Medina, and to his mother and younger brother, but also to his circle of friends and colleagues — Bruce Biggs, Pat Hohepa, Rudy Sunde, Hans Peter and Gerti Stoffels, and others.

The years passed, then one day in 1974, quite surprisingly, Andy appeared in Bratislava with his wife and sons Victor and Matthew (I am proud that the elder one bears my name). He came again in 1978.

To everyone’s surprise the communist era started drawing to a close, and in 1988 I had the opportunity to spend some two months at the University of Auckland and to travel with my friends north of Auckland as far as Te Reinga. Everything had been arranged by my friend Andy.

After the fall of the Iron Curtain in 1989 Andy arrived in Slovakia again with Medina and their younger son Matthew and we spent a couple of days together in the High Tatra, Slovakia’s only alpine mountain range, in an isolated region reminiscent of the early 20th century.

It delights me that Andy Pawley takes a real interest in Slovakia’s eventful and complicated history and in its art, as well as in the Slovak language, which is mutually intelligible with Czech, and to a considerable extent with Polish, Croatian, Serbian and even Ukrainian and Russian.

Andy’s move to Canberra in 1990 was a surprise, and several weeks spent with him and his family there were probably my last visit to the southern hemisphere. But Andy has been back to Slovakia once more. He was participating in a conference in the Netherlands and came to Bratislava en route in mid-December 2005, in the early days of a long, cruel winter. The abrupt climatic change must have been a shock to him after the Australian summer. I hope this was not his last visit.

*Kia ora, kia kaha, kia manawanui.*
A brief (incomplete) history of the Oceanic lexicon project

MEREDITH OSMOND

‘Go and see Andy Pawley. I gather he has some money for a research assistant.’ This suggestion came from Bob Dixon at the end of 1990. I had approached him for suggestions as to how I might continue to work in linguistics, having recently finished an MA and not being in a position to start a PhD. I suspect my reply was ‘Who?’ This was in the days when there was little interaction between the two Linguistics Departments, one in the Faculties, the other in the Research School of Pacific and Asian Studies. Andy had been in his job less than a year and had not yet developed the close ties that now exist between the two departments. At any rate, I went to see Andy Pawley—the first time I had set foot in the magical Coombs Building—and my life changed.

Back in Auckland, Bruce Biggs, who had supervised Andy’s PhD thesis on Kalam, had begun in the 1960s to compile Pollex, the term given to the data file of Proto Polynesian reconstructions, listing supporting data gathered from 36 Polynesian languages. In Hawaii, where Andy was also lecturing on a time-share arrangement with Auckland, Bob Blust was beginning a similar compilation of reconstructions, this time working on the entire Austronesian family, work substantially begun by the German ex-army doctor, Otto Dempwolff, back in the 1920s and 30s. Andy, good friend and colleague of both Bruce and Bob, would have understood early in his working life that the Pacific offered almost a laboratory situation for studying several hundred closely related languages, sufficiently recent in origin to permit soundly based theories of subgrouping and reconstruction of their ancestor language. A number of people had been working since the 1980s on Oceanic subgroups like South-east Solomonic (Frank Lichtenberk), Central Pacific (Paul Geraghty), North/Central Vanuatu (Ross Clark), Southern Vanuatu (John Lynch), New Caledonia (Françoise Ozanne-Rivierre) and Micronesia (Byron Bender and colleagues). At the University of Hawai’i Andy had worked on compilations of Proto Oceanic with George Grace. His arrival at the ANU in 1990 created an opportunity to join forces with Malcolm Ross, who had recently completed a vast undertaking, the subgrouping of approximately 250 Austronesian languages of north-west Melanesia, with putative POc reconstructions and subsequent sound changes for each language. It was a happy combination of interests and strengths, the one with particular expertise in Western Oceanic, the other in the East, and the two lost no time in combining forces.
The Terminologies Conference, held at the ANU in October 1990, gives some idea of the way Andy’s thinking on Proto Oceanic reconstruction was moving. Its theme was the study of Austronesian etymologies, with particular reference to terminologies likely to elucidate the culture history of the Austronesian-speaking peoples. Invited contributors provided papers on topics like the evolution of Oceanic bird names (Ross Clark), Proto Oceanic terms for food preparation (Frank Lichtenberk), early Austronesian terms for canoe parts (Andrew and Medina Pawley) and Proto Micronesian terms for the physical environment (Jeff Marck). Building on this basis, if the world as understood by Proto Oceanic speakers were explored domain by domain, it should be possible to reconstruct not only terms for items with physical existence but also terms that delved into non-material fields such as their beliefs and superstitions, their numerical system, their kinship system, and the ways in which they talked about time and space. Descriptions of particular domains in ethnographic accounts of communities either living or recent provide a likely framework in which to seek cognates for reconstructable terms. Comparative linguistics thus offers opportunities to reconstruct aspects of the unwritten past that are not recoverable through any other means.

So in early 1991, Andy and Malcolm, with me as research assistant, started out on what was shaping up as an ambitious project encompassing several volumes designed to flesh out the culture and environment of those people who lived somewhere in the Bismarck Archipelago around three and a half thousand years ago and whose language was thereafter carried to the widest reaches of the Pacific, begetting some 450 daughter languages. For the first few years, during which two volumes were produced, the work was part of our regular departmental duties. From 2003 two successive Australian Research Council grants provided resources for me to continue working on the project and for one additional person to be employed. First Bethwyn Evans and then Jeff Marck filled this role.

Initially three volumes were envisaged, Material Culture, The Physical World and People and Society. As often happens, the more information became accessible, the further we delved into topics, the more we wanted to include. The initial three now look like becoming seven. Physical World, for instance, became two volumes, then three, with the Physical Environment published in 2003, Plants in 2008 and Animals in 2010.

Different volumes create their own special spheres of exploration. Volume 1, Material Culture, involved familiarity with archaeological research. Volume 2, the Physical Environment, dealt not only with landscape and seascape but also with heavenly bodies and concepts of time and space. Volume 3, Plants and Volume 4, Animals, lent themselves to exploration of taxonomies. Volume 5, Human beings, body and mind and Volume 6 Social organisation, behaviour and cosmology, both currently in progress, are possibly our most demanding volumes, dealing with ways in which POc speakers conceptualised their abstract world. Volume 7 will contain a description of the grammatical categories of the language together with indexes of all the foregoing volumes. These three remaining volumes will take several more years, well into the retirement of all three editors.

An anonymous reviewer assessing our most recent grant application wrote that this work ‘will produce one of those rare monuments to scholarship—and to the people under study—that will still be read in a hundred years’ time’. However long it may survive as a resource, it should, as Andy envisaged it should, prove to be a comprehensive basis for future scholars carrying out research in various directions, from adding cognates which permit higher level reconstructions, to more subtle analysis of subgroups and migration.
A brief (incomplete) history of Oceanic lexicon project

patterns, to further analysis of the processes of linguistic change, and wider typological issues. And not only as a resource for linguists, but also for culture historians, archaeologists and others interested in the prehistory of the Pacific region. It will also, we hope, serve as a written history for those people who seek a wider understanding of their own past.
‘Hey brah! How you stay?’

It was 1983. We were in Lihue, the big smoke of Kauai, and we had just been introduced to a young man around our own age (14). ‘Oh at my grandma’s,’ Victor replied. There followed a long pause as both parties, still smiling at each other, tried to work out exactly what was communicated. It dawned on us later that we had been asked ‘How are you doing?’, not ‘Where am I living?’ Yeah what a lolo head.

Such is the curse of growing up with Andrew Kenneth Pawley as Dad: on one hand great exposure to other countries and cultures, and on the other a mixture of cultural confusion and disorientation. We were not the only victims. Mum enjoys visiting Europe with Dad, but she’s long given up trying to cope with the language demands. She got through France by learning to say *je voudrai* (‘I would like’), and then merely pointing a finger at whatever she wanted and hoping to hell the attendant or waiter could follow the direction.

When asked what he does, Dad likes to respond that he is a failed anthropologist. This is code for Dad having started off studying anthropology but getting sidetracked by linguistics somewhere along the way. Who exactly is to blame for this turn of events remains unclear (though we believe Bruce Biggs had a hand in it), but we can’t complain.

Dad’s discovery and pursuit of linguistics translated into lasting memories for us. By the time we had reached early double figures in age we had travelled around Micronesia and Polynesia, as well as North America and Europe. We had swum in the pristine water off the Yalobi village in Waya in western Fiji, while Dad continued his fieldwork with Timoci Sayaba. We had sat in Sproul Plaza in Berkeley, watching the best and brightest (and weirdest) of California walk to and from class, while Dad plundered the campus bookstores. We had travelled through the east cape of the North Island of New Zealand, listening to Dad speak to the locals, their faces betraying surprise as some pakeha conversed with them in Maori.

Dinner at the Pawley household was often a continuation of Dad’s teaching duties. He would regale us over a meal (which was not a meal in Dad’s eyes unless there was some kind of pudding at the end, preferably incorporating both cream and ice cream) on his or a colleague’s latest research. Most kids get to talk about sports with their Dad, and we did that too, but more often than not it was something to do with Papuan languages, Proto
Oceanic lexicons, or speech patterns and vernacular English. We would generally switch-off once Dad used the word ‘cognate’ more than twice in the space of five minutes.

We suspect many of Dad’s friends and relatives never caught on to the fact that he used them as field research. Dad has a great mate in Tasmania, Max Hodgetts. We would go on bushwalks near Max’s house in Scottsdale and Dad was ever alert to the way Max described in his various stories this tree or this piece of farming equipment. Later on, like Dr Higgins in *My Fair Lady*, he’d pass on his observations to us: ‘Notice how Max always uses “he” when talking about a tree and a “she” when talking about his truck? Fascinating!’ We trust Max gets due acknowledgement in Dad’s papers.

Even we kids were subjected to the same treatment. Dad kept journals on our speech development, a practice he has maintained with his grandkids. We are glad he did because the journals make for a tremendous and hilarious read.

One of Dad’s passions has been documenting the Wayan language. The creation process is now rivalling Charles Darwin’s *On the Origin of Species*. Like Darwin, Dad has been labouring at his prime hobby for decades. Unlike Darwin, Dad has no Alfred Wallace to spur him on and publish his masterpiece. This is a somewhat unfair statement as Dad has been busy on many fronts, not the least trying to run linguistic departments and navigate the twisting corridors of university administration while trying to teach and conduct research.

It’s fair to say that Dad is something of a workaholic. Working late into the night was (and remains) a common event, likewise working on weekends and on holidays. Some concessions are made. Dad sits in front of the TV when Test cricket is broadcast but invariably a draft chapter of some dictionary will be found on his lap. Dad looks up when a mighty ‘HOW IS HE?!’ is yelled by a bowler, but then returns to the task of revision. We sometimes think Dad is something of a junkie: he’s not happy unless he has a dictionary or draft paper in his hands to work on.

Mum, of course, has been on the front line. Coping with Dad’s idiosyncrasies is no mean feat, especially when it comes to travel and catching flights. For some reason Dad has never got a handle on laptops and mobile computing. The consequence was that he would insist on working to the last possible minute at the office, fixing up and printing out his papers and presentations. Mum would be at home, nervously imagining that the final boarding call had been made. It could get that close.

Victor even remembers Dad instructing him to speed to Canberra airport as take-off was scheduled for about 30 minutes away (they got there with ten minutes to spare).

As a reward we expect Dad to take Mum on many international trips, and to get to the airport at least an hour before departure.

We don’t know if scholars truly retire. We doubt Dad will. His mind is far too active and we can only hope this affliction endures. Besides, he still has a Wayan dictionary to finish.
Andrew Pawley’s writings

Andrew Pawley’s publications and other substantial writings are here divided into five main categories:

A. Austronesian
B. Papuan
C. General linguistic theory
D. Biographical notes, edited festschriffts
E. Regional and historical surveys and other topics

Each main category has a number of subcategories. A few ‘forthcoming’ items are included (works prepared for publication but not yet submitted) and a few unpublished papers that have been quite widely circulated. The unpublished items have an asterisk before the date of writing.

Abbreviations and publication sites of journals

AA American Anthropologist (Washington)
ARA Annual Review of Anthropology (Palo Alto)
CA Current Anthropology (Chicago)
JPH Journal of Pacific History (Routledge)
JPS Journal of the Polynesian Society (formerly Wellington, now Auckland)
Kivung Former name (1968–80) of Language and Linguistics in Melanesia (Ukarumpa PNG)
LLM Language and Linguistics in Melanesia. Journal of the Linguistic Society of Papua New Guinea (Ukarumpa) See Kivung
NZJA New Zealand Journal of Archaeology. Journal of the New Zealand Archaeological Association (University of Otago, Dunedin)
OL Oceanic Linguistics (Honolulu)
Te Reo Journal of the Linguistic Society of New Zealand (formerly Auckland, now Christchurch)
A. **AUSTRONESIAN**

A1. Polynesian languages and culture history

<table>
<thead>
<tr>
<th>Year</th>
<th>Title and Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1966</td>
<td><em>Samoan phrase structure: the morphology-syntax of a Western Polynesian language</em>. Bloomington, Indiana University Archives of Languages of the World. pp.63. (= <em>Anthropological Linguistics</em> vol. 8, No. 5)</td>
</tr>
<tr>
<td>1967</td>
<td>The relationships of the Polynesian outlier languages <em>JPS</em> 76:259–296.</td>
</tr>
</tbody>
</table>
The history of research on the Samoan language. To appear in *Journal of Samoan Studies*.

**A2. Fijian and Rotuman**

- **1970** Grammatical reconstruction and change in Polynesia and Fiji. See A1.
A3. Austronesian languages of New Guinea


A4. Wider Austronesia linguistics and culture history: sole and joint papers


1998 (Roger Green and Andrew Pawley) Architectural forms and settlement patterns. In M. Ross, A. Pawley and M. Osmond, eds The lexicon of Proto Oceanic. The culture and environment of ancestral Oceanic society: vol. 1 Material culture, 37–65. [Abridged and slightly revised version of the following item]


2010 Retention and replacement of Proto Oceanic basic vocabulary in the Central Pacific languages. Paper read at 8th International Conference on Oceanic Linguistics, Auckland, January 2010.
<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Contributors</th>
<th>Details</th>
</tr>
</thead>
</table>

**A5. Austronesian linguistics: books, encyclopaedia articles, surveys and edited volumes other than festchrifts**

<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Contributors</th>
<th>Details</th>
</tr>
</thead>
</table>
B. PAPUAN (NON-AUSTRONESIAN) LANGUAGES AND CULTURES OF NEW GUINEA

B1. Kalam language and culture


Andrew Pawley’s writings


2010 (Juliette Blevins and Andrew Pawley). Typological implications of Kalam predictable vowels. *Phonology*.


*in press* *Helter skelter* and ſugl ſagl: English and Kalam rhyming jingles and the psychic unity of mankind. In Ger Reesink and Ken McElhanon, eds *A mosaic of language and culture: celebrations of the academic career of Karl Franklin*. (?) Dallas: Summer Institute of Linguistics.

B2. Papuan comparative historical and comparative typological

1995 C.L. Voorhoeve and the Trans New Guinea Phylum hypothesis. In Connie Baak, Mary Bakker and Dick van der Meij, eds *Tales from a concave world. liber amoricum Bert Voorhoeve*, 83–122. Leiden University, Projects Division, Department of Languages and Cultures of South-East Asia and Oceania.


B3. Edited books and reviews


2007 Various short contributions to I.S. Majnep and R. Bulmer, *Animals the ancestors hunted: an account of the wild mammals of the Kalam Area, Papua New Guinea* (see above).

(a) Editors’ preface, pp. xix–xxxvii.

(b) Part IV. Madaw, the ground cuscus, pp.49–52.

(c) (Robin Hide and Andrew Pawley) Part V: Women’s prime game the bandicoots, pp.173–5.

(d) (Andrew Pawley and Robin Hide) Part VI: The spotted quoll; and the water-rats and water-side rat, pp.190–192.

(e) (Robin Hide and Andrew Pawley) Part VII: The most important as mammals: The colonial-nesting small rodents and some other small creatures of the forest, pp.203–209.
C. GENERAL LINGUISTIC THEORY, FORMULAIC LANGUAGE, LEXICOGRAPHY, AND ENGLISH LANGUAGE

Category C consists of papers that are mainly to do with general linguistic theory, lexicography and varieties or genres of English. A few items that have a theoretical focus and were cited under A and B are cross-listed here.

C1. Idiomaticity, speech formulas, lexicalisation, dictionary-making


*1972-87 (Frances Syder and Andrew Pawley) Transcripts of conversations (in English, from Tasmania and New Zealand). Approx. 300,000 words.


*1982 (Andrew Pawley and Frances Syder) The making of an oral literature. [= Chapters 11 and 12 of The spoken language (14 units for schools, typescript. pp.90) dealing with story-telling in rural Tasmania.]


1985 On speech formulas and linguistic competence. Lenguas Modernas (Chile) 12:84–104.


C2.  Other studies in grammar and discourse


*1972-87  (Frances Syder and Andrew Pawley). Transcripts of conversations (in English) from Tasmania and New Zealand. Approx. 300,000 words.


*1982  (Andrew Pawley and Frances Syder) The making of an oral literature. [= Chapters 11 and 12 of *The Spoken Language* (14 units for schools, typescript. pp.90) dealing with story-telling in rural Tasmania.]
1987  Encoding events in Kalam and English: different logics for reporting experience. See B1.


1993  A language which defies description by ordinary means. See B2.


C3. Varieties of English and Tok Pisin


D. BIOGRAPHICAL NOTES, FESTCHRIFTS AND COMMEMORATIVE VOLUMES


E. MISCELLANEOUS: GENERAL SURVEYS AND OTHER TOPICS


~ Prologue ~
8

A metrical system that defies description by ordinary means

ALAN RUMSEY

One of the first times that I met Andy Pawley — and probably the first time that my wife and co-worker Francesca Merlan and I met him together — was in Brisbane in May of 1981 at a meeting of the linguistics section of the Australia and New Zealand Association for the Advancement of Science. We had stopped off for that meeting while on the way to Papua New Guinea for our first fieldwork there. We had planned to do that work somewhere in the Nebilyer Valley west of Mount Hagen, but had little idea what we would find there. Andy was full of sound advice based on his already long experience in the Kalam region some 100 km to the northeast, and with his help and others’ we were able to get started within the next few months on what has turned out to a long series of research projects based in the Ku Waru (‘Steep Stone’) dialect area which straddles the Tambul Range along the western edge of the Nebilyer Valley and north-eastern reaches of the Kaugel Valley on the other side of the range.¹

After returning from our first extended fieldwork there Francesca and I continued to benefit from Andy’s work, especially from his insights into the grammar of Papuan languages. As many readers of this volume will know, over a quarter-century ago Andy wrote a paper which became famous among Papuanists long before it was published, entitled ‘A language that defies description by ordinary means’ (Pawley 1993). The language was Kalam — a typical Highland New Guinea language in most respects, but one which in those same respects presented a radical challenge to linguists’ understanding of what is ordinary, especially with regard to verbal semantics, and the question of where the grammar stops and the lexicon begins. Here I wish to honour Andy’s work by considering some aspects of a Ku Waru genre of verbal art which seems to present a similar challenge to current linguistic theory in another domain, namely with respect to the nature of poetic

¹ For their helpful comments on earlier versions of this paper I would like to thank Francesca Merlan and Don Niles. I am grateful to Don Niles also for his extensive research collaboration with me on tom yaya kange, for sharing the results of his insightful work on their musical aspects, and for allowing me to reproduce his musical transcription in Figure 4. Many thanks to Paul Kiparsky for his sponsorship on a period of a sabbatical leave that I spent in 2005 in the Department of Linguistics at Stanford University, and for inviting me to participate in a series of workshops on Language and Poetic Form at Stanford, including a day-long ‘Poetics Fest’ on 14 May, 2005, at which I gave a presentation from which the present paper has arisen. Thanks also to Kristin Hanson and others who discussed that presentation with me. For funding my research on tom yaya kange and related matters since 1981, I acknowledge the Australian Research Council, the University of Sydney and The Australian National University.
meter. The genre in question is called tom yaya kange. This is one of a number of related genres of sung narrative that are found across a large area of the Papua New Guinea Highlands. Since 2003 these genres have been the subject of an interdisciplinary research project based at The Australian National University.\(^2\) Drawing on the results of my collaboration on that project with musicologist Don Niles, here I will be concerned with the rhythmic properties of tom yaya kange; with the question of whether, and in what ways it may be described as a metrical genre; with how its prosodic properties are related to those of ordinary speech; and with the implications of its metrical system for linguistic theory. As necessary background to that discussion, I will begin with a summary of some aspects of current theories of poetic meter. This will be followed by a brief overall introduction to tom yaya kange.

1 The nature of poetic meter

Metrical poetic systems are ones which create special aesthetic effects by establishing repeating rhythms which are more regular than those of most ordinary speech. Consider, for example, the following, opening line from Henry Wadsworth Longfellow’s poem Evangeline:

This is the forest primeval. The murmuring pines and the hemlocks,

The words in this line are specially selected and put together by the poet in such a way as to give to each line a regular rhythm with six relatively evenly-spaced beats, which establish a grouping of the syllables as follows:

This is the forest primeval. The murmuring pines and the hemlocks

Each of these groupings of syllables is what is known to poetics as a foot. In this case each foot except the last one in the line consists of one stressed syllable followed by two unstressed ones. This is actually a fairly unusual kind of foot within English poetry, which more commonly makes use of so-called ‘rising’ rhythms, with a strong beat on the last syllable of the foot rather than the first. There is a special reason for Longfellow’s use of a ‘falling’ foot here: the metrical line of Evangeline is deliberately modelled on that which is used throughout the Iliad and Odyssey of Homer (and by Virgil and other Latin poets who also took Homer as a model). Though perhaps not a stellar piece of English verse in its own right, as a piece of imitation-Homeric prosody Evangeline is a tour de force, not only for its rather forced neo-classical dactylic hexameter rhythms, but for the way in which Longfellow manages to overcome an important difference between Homeric Greek prosody and that of modern English, namely, the fact that Homer’s metrical line is not built out of an alternation between stressed and unstressed syllables (as English poetry is). Rather, it is based on an alternation between long syllables and short ones, the long ones being, roughly speaking, those with long vowels or diphthongs and/or with more than one consonant following the vowel, and the short ones being those without either.\(^3\) The use that Homer makes of that kind of difference is illustrated by the following display of the first line of The Odyssey:

---

\(^2\) For details of this project see http://rspas.anu.edu.au/anthropology/chanedtales/.

\(^3\) For a fuller specification of this which takes account of certain other factors, see Pharr (1959:210, 339–342).
A metrical system that defies description by ordinary means

Andra moi ennepe mousa pollytropon hos mala polla
‘Tell me, oh Muse, of that man of many turns, who in many ways…’

As can be seen, each foot begins with a long syllable. The first five feet end with two short syllables and the final foot with a single short syllable. This dactylic hexameter line type is used through the entire Iliad and Odyssey. As can be seen by comparing this line from The Odyssey with the one from Evangeline above, Longfellow has created an English line which precisely replicates the Homeric one, with the difference that in place of Homer’s long and short syllables Longfellow uses stressed and unstressed ones respectively.

The point of this excursus on Homer and Longfellow is not to impart a literary flavour to the discussion, but to show how poets working in languages with very different prosodic systems can nonetheless produce similar metrical structures — indeed ones which, at a certain level of abstraction may be regarded as identical. That identification can be made by subsuming the Greek opposition between long and short syllables and the English opposition between (primary-) stressed and relatively less stressed ones under a single more general opposition between ‘strong’ syllables and ‘weak’ ones. That is exactly what is done within most of the current universalising frameworks for the comparative study of poetic meter, including what is perhaps the most rigorous and well established of them, that of Kristin Hanson and Paul Kiparsky (1996). Building on earlier work on English prosody by Halle and Keyser (1971) and Kiparsky (1977), Hanson and Kiparsky (1996:289) assume that

the basic form of any periodic meter comprises a fixed number of headed binary FEET, where the essential character of a head is rhythmic prominence. Each foot thus consists of two METRICAL POSITIONS; for convenience we label the head STRONG (S) and the nonhead WEAK (W).

The units which occupy these metrical positions are independently specified within a given language as relatively strong or weak — for example, in English as (groupings of) stressed and unstressed syllables or in Homeric Greek as long or short ones. Given the inherent specifications of each syllable of every word for relative strength or weakness, and the fact that the line structures of metrical verse establish their own patterns of periodic alternation among strong and weak syllables, it follows that there are strong constraints on the words that can occupy any given position within the line, and that the poet’s skills must include an ability to operate creatively within those constraints.

2 Tom yaya kange

In common with all the other regional genres of Highlands sung narrative that have been studied so far, Tom yaya kange are composed and performed by only a small minority of people — mostly men but some women also — who are recognised as having special skills that allow them to do so. In this respect the genre contrasts strongly with all of the recognised song and story genres in the region, which are performed by nearly all adults.

---

4 Not every line in Homer has exactly the same distribution of short syllables as the one shown here. While every foot always begins with a long syllable, the foot is sometimes completed by a single long syllable rather than two short ones. The final (sixth) foot of the line may be completed either by a single short syllable as in this example, or by a single long syllable.

5 For details of these see the map and bibliography at: http://rspas.anu.edu.au/anthropology/chantedtales/.
and older children. In particular, the latter is true of the story genre called *kange,* of which *tom yaya kange* are the sung version. While nearly any of the *kange* can be adapted for *tom yaya* — style sung delivery, in practice only a restricted range of them tend to be, and, conversely, the most commonly-performed *tom yaya kange* draw on a stock of plots and characters that are seldom if ever rendered in spoken form. The plots of *kange* which are performed in *tom yaya* style — i.e. the plots of *tom yaya kange* — almost always centre on a journey which is taken by a young man, usually to court a beautiful young woman that he has heard about in a distant place. In preparation for the journey he adorns himself and usually kills a pig to roast and take the pork with him to give the young lady and her family. He is usually successful in winning her hand, and they set out to return to his place and be married. In some of the stories this homeward journey is completed and the marriage takes place, while in others it is foiled and one or both of the partners fall victim to mishaps or attacks by jealous rivals.

There is much individual stylistic variation among bards, some of which will be exemplified below. But there are certain formal features that have been found to be common to the *tom yaya kange* of all the eighteen bards whose work has been examined to date. These are as follows:

1. Clear division into lines of fixed length, each ending in an added vowel *e, o* or *a*
2. Use of formulaic expressions, each comprising one or more whole lines
3. Organisation of the lines into repeating two-part melodies, the second part being a line-by-line variation on the first
4. A strong tendency for each line to comprise an integral syntactic unit
5. Extensive parallelism across lines

In addition to these five features, most but not all *tom yaya kange* display two others, always in conjunction with each other:

6. Rhythmic organisation of each line into a fixed number of beats or feet
7. A strong tendency for each beat except the final one to be associated with a single word or bound morpheme

Features 1–6 have been discussed elsewhere. Here I will focus in particular on the features that are most relevant for the comparative study of metrical systems: points 1, 3, 6 and 7. Before doing that I will first introduce some relevant details about the Ku Waru language.

### 3 Aspects of Ku Waru segmental phonology and prosody

Table 1 shows the phonemic inventory of Ku Waru and the practical orthography used in the transcripts below.

---

6 For the distinguishing features of this genre in relation to the other major story genre, called *teman,* see Merlan (1995).

A metrical system that defies description by ordinary means

Table 1: Ku Waru segmental phonology and practical orthography

<table>
<thead>
<tr>
<th>CONSONANTS</th>
<th>Labial</th>
<th>Apico-alveolar</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Stop</td>
<td>p</td>
<td>t</td>
<td>s [s–ts]</td>
<td>k</td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td>ny, yn [ŋ]</td>
<td>ng [ŋ]</td>
</tr>
<tr>
<td>Continuant</td>
<td>w</td>
<td>r</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>Retroflex tap</td>
<td>r̃t [r̃t–t̃]</td>
<td>l</td>
<td>ly, yl [ʎ]</td>
<td>1 [gl–k–l] (sometimes also spelled gl or kl)</td>
</tr>
</tbody>
</table>

VOWELS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>u</th>
<th>o</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Ku Waru, all polysyllabic words when pronounced carefully in isolation have within them a single syllable with a pitch peak which is higher than that of any other syllable within the word; for about 80% of them it is the final syllable. All monosyllabic words also have a pitch peak which is of about the same level as that of the high-toned syllable in polysyllabic words. In the practical orthography which is used here, for polysyllabic words, high pitch, if shown at all, is shown (with an acute accent mark `) only when it falls on a non-final syllable. All other polysyllabic words have the default pitch peak on the final syllable. Examples of words with contrasting pitch patterns are:

\[
\begin{align*}
koma & \rightarrow \acute{\koma} \\
\text{‘carry on one’s shoulder’} & \rightarrow \text{‘wet’}
\end{align*}
\]

\[
\begin{align*}
toba & \rightarrow \acute{\toba} \\
\text{hit:FUT.3s} & \rightarrow \text{hit:NF.3s}
\end{align*}
\]

\[
\begin{align*}
\text{‘he/she/it will hit’} & \rightarrow \text{‘he/she/it having hit’}
\end{align*}
\]

Unlike in English and other languages with stress-accent systems, the syllables with pitch peaks in Ku Waru are not, ipso facto, longer or of higher energy level than others. Regardless of pitch level, word-final syllables, especially in connected speech (as opposed to elicited single words) tend to be shorter than initial and medial ones. Indeed, when final syllables are open (i.e. vowel final) ones with high vowels, in connected speech the vowel

---

8 Almost all monosyllabic words have identical pitch patterns, in which the pitch peak generally occurs within the first half of the syllable nucleus and is followed by a fall over the rest of it. We have come across only a handful of exceptions to this, and all of them involve phonetically long vowels which are probably best interpreted as geminate vowels or bi-moraic syllables, within which one of the vowels or morae bears the high tone for the word. The most commonly occurring example is the word naá ‘not’, which has a long vowel with a gradually rising pitch over most of all of its duration, with a fall occurring near the end of it if at all. It contrasts with the word na ‘I’, which has a shorter vowel.
is often omitted altogether, leaving the word with no high-toned syllable. In such cases, the
underlying presence of a final high-toned syllable is signalled by the absence of any high-

4 Prosodic features of *tom yaya kange*

All *tom yaya kange* which have been recorded and analysed to date are delivered in
audibly distinct lines, which are regularly marked off by an added vowel *e*, *o* or *a* at the
end of the line, or, if the last word ends in one of these vowels, by a lengthening of it. In
almost all of those *tom yaya kange* (including all them which were said to be really well
composed and performed), the line is divisible into a fixed number of metrical units which
remains constant throughout the performance. Following the established usage discussed
above, I call each of those units a ‘foot’ and the associated rhythmic unit a ‘beat’. Each
foot generally⁹ comprises an integral number of syllables, ranging from one to three.

Text 1: First 16 lines of a (1070-line) *tom yaya kange* in five-beat style by Paulus Konts¹¹

<table>
<thead>
<tr>
<th>Line</th>
<th>Words</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>puku tópa lku-d urum e</td>
<td>He jumped and came into the house</td>
</tr>
<tr>
<td>2</td>
<td>dali pula manya lyirim e</td>
<td>He removed his banana leaf apron</td>
</tr>
<tr>
<td>3</td>
<td>ola pula wai lyirim e</td>
<td>And put on his cordyline kilt.</td>
</tr>
<tr>
<td>4</td>
<td>ngi kapola mari tekin</td>
<td>Well done, my lad, well done!</td>
</tr>
<tr>
<td>5</td>
<td>kanab a kélipa púpa a</td>
<td>As I watched he went on his way</td>
</tr>
<tr>
<td>6</td>
<td>koroka komunga kai kanunga a</td>
<td>Headed for Koroka Mountain.</td>
</tr>
<tr>
<td>7</td>
<td>ola púpa mólpua mel</td>
<td>He climbed to the top and stayed</td>
</tr>
<tr>
<td>8</td>
<td>tubal kop ekeda lyirim</td>
<td>With a Jew’s harp in one hand</td>
</tr>
<tr>
<td>9</td>
<td>kulaip mingiyel ekeda lyirim</td>
<td>And a bamboo flute in the other</td>
</tr>
<tr>
<td>10</td>
<td>kanab ku kéla purum e</td>
<td>As I watched he went on his way</td>
</tr>
<tr>
<td>11</td>
<td>toku nóba lkaib turum e</td>
<td>Where he smoked his tobacco and spat</td>
</tr>
<tr>
<td>12</td>
<td>toku wale pora purum e</td>
<td>Fields of tobacco plants sprouted</td>
</tr>
<tr>
<td>13</td>
<td>toku ikilya purum kanuma a</td>
<td>And the smoke that went up in the sky</td>
</tr>
<tr>
<td>14</td>
<td>waru kupa punglau nyirim a</td>
<td>Billowed like clouds round the mountain</td>
</tr>
<tr>
<td>15</td>
<td>i kapola mari tekin a</td>
<td>Well done, my lad, well done!</td>
</tr>
<tr>
<td>16</td>
<td>kanab taka taka nyiba a</td>
<td>In my mind’s eye the story unfolds</td>
</tr>
</tbody>
</table>

This style of *tom yaya kange* with five feet or beats per line has the shortest line type found
in any of the styles that have been analysed to date. Other styles have lines with six, seven,
and eight beats or feet per line.¹² Text 2 exemplifies a style with six feet per line.

---

⁹ In almost every case where the foot consists of a word with more than three syllables, one or more of the
syllables is phonetically reduced or elided, as will be exemplified in the texts below.

¹⁰ This style was made popular, beginning in the 1980s, by a performer named Paul Pepa, through
broadcasts of his recorded performances by Radio Western Highlands, the Mt. Hagen-based local service

¹¹ The Appendix below provides interlinear glosses for Texts 1 and 2.

¹² For an analysis of an eight-beat style, see Niles (2007).
A metrical system that defies description by ordinary means

### Text 2: First 10 lines of a (363-line) *tom yaya kange* in six-beat style by Kopia Noma

<table>
<thead>
<tr>
<th>Line</th>
<th>Lines</th>
<th>Stanzas</th>
<th>Foot Structure</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kang</td>
<td>mel</td>
<td>we mel kaniyl e</td>
<td>Though the tiniest slip of a lad</td>
</tr>
<tr>
<td>2</td>
<td>kang</td>
<td>mai</td>
<td>pup yaka nyirim e</td>
<td>That boy strode from perch to perch.</td>
</tr>
<tr>
<td>3</td>
<td>kang</td>
<td>komunga</td>
<td>mong yaka nyirim e</td>
<td>That boy strode from mountain to mountain.</td>
</tr>
<tr>
<td>4</td>
<td>ukuni</td>
<td>yahu</td>
<td>tobu midi nyirim e</td>
<td>He wanted to slay the Ukuni.</td>
</tr>
<tr>
<td>5</td>
<td>kobulka</td>
<td>yahu</td>
<td>tobu midi nyirim</td>
<td>He wanted to slay the Kobulka</td>
</tr>
<tr>
<td>6</td>
<td>kang</td>
<td>mel</td>
<td>we mel kaniyl e</td>
<td>Though the tiniest slip of lad</td>
</tr>
<tr>
<td>7</td>
<td>kang</td>
<td>pidi-</td>
<td>yap mel kaniyl e</td>
<td>Who’d been ignored since the day he was born</td>
</tr>
<tr>
<td>8</td>
<td>pilyini</td>
<td>kub</td>
<td>nai ko, nyirim e</td>
<td>And who’s ever heard such a tale?</td>
</tr>
<tr>
<td>9</td>
<td>kanuni</td>
<td>kub</td>
<td>nai ko nyirim e</td>
<td>And who’s ever seen such a thing?</td>
</tr>
<tr>
<td>10</td>
<td>kang</td>
<td>mai</td>
<td>pup yaka nyirim</td>
<td>That boy strode from perch to perch.</td>
</tr>
</tbody>
</table>

In these and most other *tom yaya* styles, the line-final vowel by itself comprises a single foot. As can be seen, in Noma’s six-beat style, this vowel is omitted just when Noma takes a breath at the end of the line. In Konts’ five-beat style (see Text 1) he sometimes also does this, as in lines 4 and 8, but sometimes he leaves in the line-final vowel — albeit usually shortened — and breathes after it, as in lines 12 and 16. As shown by both texts, there is a strong tendency for each of the other feet in the line besides the line-final one to consist of a single word. There are only two exceptions. One of them is in the second foot of line 5 in Text 1, where Konts adds a second semantically empty vowel *a* to fill out the line, which has only three words in it. The other exception is in the second and third foot of line 7 in Text 2, where a single word, *pidiyap* (‘neglected’) is stretched across two feet to fill out the line in lieu of a fifth word. Taking these two exceptions into account, we can say that of the 113 word tokens in both texts, 112 of them or approximately 99% occupy a single foot. Conversely, if we discount the final foot in each line, which is regularly occupied by a vocable and/or a breath (invariably in this sample), we can say that of the other 114 foot-positions in these two texts, 111 or approximately 97% of them are occupied by a single word. By contrast, of the 37 word-tokens in the first five lines of *The Odyssey*, only six, or approximately 16%, occupy a single foot, and of the 30 foot-positions in those lines, only six, or 20%, are occupied by a single word. There is a similarly low degree of correspondence between the foot and the word in Longfellow’s *Evangeline*. These examples are indicative in that: (1) the high degree of word-to-foot correspondence in Texts 1 and 2 is similar to what is found across

---

13 In some previous publications of mine (Rumsey 2001:205, 211, 228; 2005:47, 51 and 2006:339, 340) this word has been incorrectly rendered as *piditap* or *piditab*. This was due to a typographical error introduced when this text was first typed up from my hand-written field notes.

14 For the full data sample on which these counts are based — a display of the first five lines of *The Odyssey* divided up into feet — see Rumsey (2001:195). The general pattern can be seen from the first line alone, as shown above.

15 Here are the first five lines of that poem, from which interested readers can do their own tabulation:

- This is the forest primeval. The murmuring pines and the hemlocks,
- Bearded with moss, and in garments green, indistinct in the twilight,
- Stand like Druids of eld, with voices sad and prophetic,
- Stand like harpers hoar, with beards that rest on their bosoms.
- Loud from its rocky caverns, the deep-voiced neighboring ocean
the full range of metrical *tom yaya kange* performances that have been recorded and analysed; and (2) the much lower degree of word-to-foot correspondence in *The Odyssey* and *Evangeline* is typical of what is found, not only in Homer and Longfellow, but more generally in metrical verse composed within stress-based prosodic systems such as that of English, or quantity-based systems such as that of Homeric Greek.\(^{16}\)

Another difference between the *tom yaya kange* metrical system and the others discussed above is that, as far as I have been able to discover, the Ku Waru metrical line does not impose any constraints on what words may occupy a given position within the line (as opposed to the more general constraint that each foot should be occupied by a word). In particular, the *tom yaya* line does not have an internal structure based on a distinction between strong and weak syllables,\(^{17}\) nor is any role played by a distinction of that kind among syllables within words. It is relevant here to note once again that the Ku Waru tone system does establish inherent differences among the syllables of all polysyllabic words, by assigning a pitch peak to just one of them, but that, as we have seen, that high-toned syllable could not readily be treated as a ‘strong’ one. Indeed, in the great majority of Ku Waru words, where the high-toned syllable is the final one with a high vowel, that is the weakest syllable in terms of length and energy level, often disappearing completely in connected speech.

Now given the two facts about the *tom yaya* line that I have just described — the realisation of foot as word and nearly complete freedom as to choice of word to fill a given foot — how are these two facts related to each other? Minimally, they are related in that the latter is an enabling condition of the former. That is, because nearly any word can fit into any foot without having to accommodate some of its syllables within an adjacent one (as for example in the second foot of the line from *Evangeline* above, or the third foot of the line from *Homer*), it is possible to operationalise a preference for each foot to be filled by a full word.

Indeed, it may be more appropriate to speak of the foot-to-word relationship in *tom yaya kange* not just as a preferential one but as essential to the identity of the foot itself. For in the absence of audible periodic alternation between strong and weak syllables such as one finds in Homeric or Middle and Modern English verse, one of the main audible cues for the division of the *tom yaya* line into a recurring number of feet is the strong tendency for the pronunciation of each syllable within a word to be stretched or contracted in proportion to the number of syllables in the word. This is illustrated by Figure 1, which shows the line length and syllable timing within the first eight lines of Konts’ *tom yaya kange* as transcribed in Text 1; and Figure 2, which shows them for the ten lines of Noma’s *tom yaya kange* as transcribed in Text 2. While the pronunciation of the bi- and trisyllabic words is of longer duration than that of the monosyllabic words, it is nowhere near being two and three times as long respectively. Also, there is a tendency for the duration of all the words to be evened out so as to make it roughly similar to that of the added vowel which comprises the last foot of every line.

---

16 For further examples, compare the many systems of this kind that are analysed in Hanson and Kiparsky (1996) and in the various chapters in Kiparsky and Youmans (1989).

17 In a previous publication (Rumsey 2001) I described the polysyllabic foot of metrical *tom yaya kange* as being internally differentiated in this respect, with a strong tendency for the first syllable of the foot to be of longer duration than the other ones. In light of the further work I have done since then on the phonetics of *tom yaya kange* using the computer program Praat (as in Figures 1 and 2), I no longer believe that to be the case.
Figure 1: Line length and syllable timing within the first eight lines of Konti's 'tom yaya kange'
Figure 2: Line length and syllable timing within the first ten lines of Noma’s tom yaya kange, lines 1–5
A metrical system that defies description by ordinary means

Figure 2a: Line length and syllable timing within the first ten lines of Noma’s *tom yaya kange*, lines 6–10
The structure of the two kinds of *tom yaya kange* poetic line that I have just described would not seem to be readily accommodated by Hanson and Kiparsky’s theory of poetic meter, or any other account that presupposes a distinction between strong and weak prosodic units or metrical positions as a universal basis of metrical systems. For no such distinction plays a part in the metrical structure of these lines, which takes the word per se as its main linguistic building block. To a certain extent, Hansen and Kiparsky’s theory does allow for the possibility of this kind of system, in that its ‘position size parameter’, which ‘defines the maximal size of a metrical position in terms of the constituents in the universal prosodic hierarchy’, includes among possibly relevant units not only the mora, the syllable and the foot, but also, at the maximal level ‘the phonological word’ (Hansen and Kiparsky 1996:289). Hansen and Kiparsky do not discuss any examples of systems where the phonological word functions in that way, but refer the reader to Russom’s (1987) analysis of Beowulf’s Old English meter as one such system. Against a common approach which treats the lines of Old English verse as consisting of half-lines, each with two main stresses and an indeterminate number of weaker stresses, Russom (ibid.) develops an analysis in which word boundaries figure directly in the definition of metrical units. Russom points out that he is not the first to do so, citing as other precedents the concept of ‘word foot’ used by Travis (1973) in his analysis of Old Irish poetry, and a similar primacy that Roman Jakobson gave to the word in his 1938 analysis of the work of the Czech poet K.H. Mácha (Jakobson 1979a).

In Jakobson’s 1979 ‘Retrospect’ on that analysis and the other papers on poetics which are brought together in volume 5 of his Selected Writings, Jakobson argues strongly against the idea that all of the properties which figure in systems of poetic meter must be audible features. In Jakobson’s view this assumes too close a parallel between poetry and music. Citing Cooper and Meyer’s (1960) point that in music, rhythmic grouping ‘on all architectonic levels is a product of similarity and difference, proximity and separation of the sounds perceived by the senses and organised by the mind’, Jakobson contrasts this with poetry, in that there:

> the fundamental grouping is the division of the verbal flow by word boundaries. This phrasing imposed on the verse by the rules of grammar is based upon the speakers-listeners’ compulsory attention to the affiliation of syllables appertaining to the same word units and to the severance of syllables belonging to separate units. The awareness of such ‘bridges’ and ‘breaks’ remains valid independently of the question, whether the physical implementation of a given boundary is obligatory or merely optional. (Jakobson 1979b:576)

In this view it is the meaningful and distributional properties of words rather than their phonological ones per se that have primacy in establishing the metrical patterns of poetry. From the discussion above it can be seen that the status of the word within *tom yaya kange* is in close accord with Jakobson’s understanding of the matter in that there is an almost total identification between word and foot. As we have seen, there is an audible manifestation of this division into word-feet, insofar as the syllables of each word tend to be lengthened or shortened as necessary to make the words of roughly equal duration. There is another audible aspect of *tom yaya kange* that also contributes to the psychological reality of the word-foot, and it is one which, notwithstanding Jakobson’s well-founded point about the differences between music and poetry, arises in this case from an interaction between the two. To show how it does, I will first describe and exemplify some of the musical properties of the genre.

---

18 I wish to thank Paul Kiparsky for pointing this out to me.
5 Musical features of *tom yaya kange*

In addition to the prosodic patterning discussed above, *tom yaya kange* also have a musical structure which is closely related to it at the level of the line and the foot. At least in the work of the performers being considered here, Kopia Noma and Paulus Konts, each foot is associated usually with a single musical pitch within a set pattern of pitch movements across the line. At a higher level of tonal organisation, these pitch movements are organised into repeating melodies, each comprising an even number of lines (or, in musical terms, ‘measures’). In the *tom yaya kange* by Konts discussed above, there is an eight-line melody which uses five pitch levels, corresponding roughly to the first six notes of the western musical scale. This melody is represented in Figure 3.

\[
\begin{array}{cccccc}
1 & 1 & 1 & 5 & 5 \\
5 & 5 & 6 & 5 & 2 \\
5 & 5 & 3 & 4 & 4 \\
2 & 2 & 3 & 2 & 2 \\
2 & 2 & 2 & 5 & 5 \\
6 & 5 & 5 & 4 & 1 \\
3 & 3 & 2 & 3 & 3 \\
1 & 2 & 2 & 1 & 1 \\
\end{array}
\]

**Figure 3:** Basic form of 8-line *tom yaya kange* melody shown in cipher notation

The first line of numbers in Figure 3 shows the musical pitch levels which map onto the corresponding five feet in line 1 of Text 1, the second line to those of line 2 in that text, etc. Since the melody is repeated, its first line also corresponds to the pitches in line 9, the second line to line 10, and so forth, in repeating eight-line cycles throughout the performance. Note that the pitch levels shown by numbers in this display apply to entire feet, regardless of the number of syllables in them, so that in the case of polysyllabic feet, all of the syllables are usually sung at roughly the same pitch level.

The same is true of Kopia Noma’s six-beat style of *tom yaya kange*. This is shown by Figure 4, which is a rough musical transcription by Don Niles of the first ten lines of the same performance that is transcribed in Text 2. Unlike the cipher notation that is used to represent the general form of Konts’ eight line melody in Figure 3, in this transcription of ten actual lines of texted *tom yaya kange*, one can see what happens to the melody when there is more than one syllable to the foot. Don Niles wishes it to be noted that the time- and pitch-values in this transcription are approximate only. Taken in combination with the more precise display of syllable timing in Figure 2, they may be taken as confirming the point that the syllables tend to be lengthened or shortened so as to even up the durations of words, whether monosyllabic or polysyllabic, thereby contributing to the aural perceptibility of the foot.

\[\text{The main difference is that the third tone of Konts’ scale is often lower than a western major third, approaching a level more like that of a minor third.}\]
Figure 4: Musical transcription of the first ten lines of *tom yaya kange* by Kopia Noma

More importantly with respect to the issue at hand, Niles’ musical transcription of Noma’s performance bears out the same point that was made above about the patterning of Konts’ musical pitches, namely, that each foot tends to be associated with a single one of them. In the case of polysyllabic feet — which correspond in the musical transcription (Figure 4) to groups of quavers and semiquavers — it is true that 13 out of the 31 of them involve some pitch movement within the group (usually by only one note). However, apart from the line-final vowels, among all of the other 23 monosyllabic feet in these ten lines, there is not a single case of what in musical terms is known as ‘melisma’, where a single syllable is stretched over two or more notes which are sung at different pitches. This happens only with the lexically empty line-terminating vowel in lines 1, 2, 4, 6, 7 and 9. Taking the monosyllabic feet together with the polysyllabic ones, we can say that, of the 58 sung feet in this ten-line sample, 45 of them, or approximately 78%, involve an association between a foot and a single musical pitch. If we exclude the eight sung feet which are not words, i.e., those filled by line-final vowels, the proportion increases to 43/50 or 86%.

---

20 I stipulate ‘sung feet’ in order to exclude the final foot of lines 5 and 10, which are each filled with a breath.
6 Conclusion

From the above it can be seen that the musical structure of metrical *tom yaya kange* is in close accord with their poetic form, both of which make use of word-length units as a fundamental building block. Heard in combination with the syllable length adjustments discussed above, the organisation of the *tom yaya* melody into foot-length, pitch-based musical units provides an additional aural cue both to the presence of the foot as a recurring rhythmic unit\(^{21}\) and to its close association with the word. In the absence of any deployable distinction between strong and weak syllables in the prosodic system of the Ku Waru language, neither the musical organisation of *tom yaya kange* nor the linguistic-metrical one makes use of any such distinction,\(^{22}\) instead treating the word itself as the basic unit. In that respect the metrics of *tom yaya kange* would seem to ‘defy description by ordinary means’. However, just as Andy Pawley’s work on Kalam has pointed the way to a revised understanding of what should be taken as ordinary with respect to the relation between grammar and lexicon, perhaps closer attention to word-based metrical systems such as the *tom yaya* one might lead to new understandings of relationships between poetry and music, and of the range of occurring or possible systems of poetic meter.

Appendix

Text 1a: Same lines as in Text 1 above with interlinear glosses

<table>
<thead>
<tr>
<th></th>
<th>puku jump</th>
<th>tópa do:NF.3s</th>
<th>lku-d house-DAT</th>
<th>urum come:PF.3s</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>e</td>
</tr>
<tr>
<td>2</td>
<td>dali banana leaf</td>
<td>pula cordyline</td>
<td>manya down</td>
<td>lyirim take:PF.3s</td>
<td>e</td>
</tr>
<tr>
<td>3</td>
<td>ola up</td>
<td>pula</td>
<td>wal down</td>
<td>lyirim take:PF.3s</td>
<td>e</td>
</tr>
<tr>
<td>4</td>
<td>nig that</td>
<td>kapola good</td>
<td>mari things</td>
<td>tekin do:PPR.2s</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>kanab see:OPT.1s</td>
<td>a</td>
<td>kélipa leave:PF.3s</td>
<td>púpa go:NF.3s</td>
<td>a</td>
</tr>
<tr>
<td>6</td>
<td>koroka Koro (name)</td>
<td>komunga mountain</td>
<td>kai good</td>
<td>kanunga a</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ola up</td>
<td>púpa do:NF.3s</td>
<td>mólpupa stay:NF.3s</td>
<td>mel like</td>
<td>e</td>
</tr>
<tr>
<td>8</td>
<td>tubal Jews harp</td>
<td>kop knotted</td>
<td>ekeda on a side</td>
<td>lyirim take:PF.3s</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>kulaip flute</td>
<td>mingiyl tube</td>
<td>ekeda on a side</td>
<td>lyirim take:PF.3s</td>
<td>e</td>
</tr>
</tbody>
</table>

\(^{21}\) Again, I owe this point to Paul Kiparsky, who offered it in response to a presentation I made on *tom yaya kange* at a workshop on Language and Poetic Form at Stanford University in May 2005.

\(^{22}\) This places them outside the compass not only of Hanson and Kiparsky’s (1996) theory of poetic meter, but also of Halle and Lerdal’s (1993) related musical-cum-linguistic model of ‘text setting’, which tries to account for the mapping of the words of any given text in any language onto the notes of a given melody. Like Hanson and Kiparsky’s model, Halle and Lerdal’s assumes a differentiation of strong and weak structural positions and syllables as its basis.
<table>
<thead>
<tr>
<th>Line</th>
<th>English</th>
<th>Amharic</th>
<th>Part of Speech</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>kanab</td>
<td>ku</td>
<td>and</td>
<td>leave</td>
</tr>
<tr>
<td></td>
<td>see:OPT.1s</td>
<td>kéla</td>
<td>go:PF.3s</td>
<td>e</td>
</tr>
<tr>
<td>11</td>
<td>toku</td>
<td>nóba</td>
<td>consume:NF.3s</td>
<td>lkaib</td>
</tr>
<tr>
<td></td>
<td>tobacco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>toku</td>
<td>wale</td>
<td>pora</td>
<td>purum</td>
</tr>
<tr>
<td></td>
<td>tobacco</td>
<td>sprout</td>
<td>completely</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>toku</td>
<td>ikilya</td>
<td>purum</td>
<td>kanuma</td>
</tr>
<tr>
<td></td>
<td>tobacco</td>
<td>smoke</td>
<td>go:FPST.3s</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>waru</td>
<td>kupa</td>
<td>punglau</td>
<td>nyirim</td>
</tr>
<tr>
<td></td>
<td>mountain side</td>
<td>billow</td>
<td>move</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>i</td>
<td>kapola</td>
<td>mari</td>
<td>tekin</td>
</tr>
<tr>
<td></td>
<td>that</td>
<td>good</td>
<td>things</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>kanab</td>
<td>taka</td>
<td>taka</td>
<td>nyiba</td>
</tr>
<tr>
<td></td>
<td>see:OPT.1s</td>
<td>quietly</td>
<td>quietly</td>
<td></td>
</tr>
</tbody>
</table>

Text 2a: Same lines as in Text 2 above with interlinear glosses

<table>
<thead>
<tr>
<th>Line</th>
<th>English</th>
<th>Amharic</th>
<th>Part of Speech</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kang</td>
<td>mel</td>
<td>we</td>
<td>mel</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>small</td>
<td>very small</td>
<td>that</td>
</tr>
<tr>
<td>2</td>
<td>kang</td>
<td>mai</td>
<td>earth</td>
<td>pup</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>earth</td>
<td>resting place</td>
<td>stride</td>
</tr>
<tr>
<td>3</td>
<td>kang</td>
<td>komunga</td>
<td>mong</td>
<td>yaka</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>mountain</td>
<td>point</td>
<td>stride</td>
</tr>
<tr>
<td>4</td>
<td>ukuni</td>
<td>yabu</td>
<td>people</td>
<td>tobu</td>
</tr>
<tr>
<td></td>
<td>(tribe name)</td>
<td>hit/kill:FUT.3s</td>
<td>only</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>kobulka</td>
<td>yabu</td>
<td>people</td>
<td>tobu</td>
</tr>
<tr>
<td></td>
<td>(tribe name)</td>
<td>hit/kill:FUT.3s</td>
<td>only</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>kang</td>
<td>mel</td>
<td>we</td>
<td>mel</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>small</td>
<td>very small</td>
<td>that</td>
</tr>
<tr>
<td>7</td>
<td>kang</td>
<td>pidi-neglected</td>
<td>yap</td>
<td>mel</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>neglected</td>
<td>like</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>pilyini</td>
<td>kub</td>
<td>nai</td>
<td>ko</td>
</tr>
<tr>
<td></td>
<td>hear:FUT.2s</td>
<td>and</td>
<td>who</td>
<td>ko,</td>
</tr>
<tr>
<td>9</td>
<td>kanuni</td>
<td>kub</td>
<td>nai</td>
<td>ko</td>
</tr>
<tr>
<td></td>
<td>see:FUT.3s</td>
<td>and</td>
<td>who</td>
<td>ko</td>
</tr>
<tr>
<td>10</td>
<td>kang</td>
<td>mai</td>
<td>pup</td>
<td>yaka</td>
</tr>
<tr>
<td></td>
<td>boy</td>
<td>earth</td>
<td>resting place</td>
<td>stride</td>
</tr>
</tbody>
</table>
References


~ Across the disciplines ~
Genes, languages and agriculture in New Guinea population history

ROBERT ATTENBOROUGH

Genes and languages: do their histories mirror each other?

We humans are both an evolved and a cultural species. Most individuals amongst us, most of the time, are likely to learn, from one or both of our biological parents, a first language which was also theirs, and which in turn will probably be our children’s. Even where children have grown up with considerable sociolinguistic complexity, such as multilingualism or linguistic exogamy, as frequently is the case in small scale societies, they still most often become native speakers of at least one language shared with at least one of their parents. Can we extrapolate from this that there is sufficient parallelism in the transmission of genes and languages that we should be able to see them mirror each other on a larger scale?

If it is true that genes and languages most often travel together, down the generations and across the landscape of any particular region, the predicted outcome of population history in the long run would be patterns in which they co-occur, maybe even strongly enough that we can use evidence about one in drawing inferences about the other. Such co-occurrence might be manifested, for example, as spatial coincidence of relatively sharp population differentiation in genetic terms with linguistic boundaries. There will, of course, be instances in which the historical circumstances are too fraught for parallel genetic and linguistic transmission to be a realistic expectation. But if it holds most of the time, that may be often enough to generate interesting consequences. In this essay, motivated by an interest in exploring that parallelism in the case of New Guinea but drawing ideas also from the case of Europe, I should like to begin by asking what happens when we look for associations between patterns of genetic and linguistic resemblance and boundary, and attempt to interpret them in terms of population pasts.

1 I thank the editors for the invitation to contribute to this volume; and Peter Bellwood, Malcolm Ross and an anonymous reviewer for their constructive comments which have helped to improve this paper.

2 I use ‘gene’ here as a shorthand for any segment of DNA sequence: non-coding and non-genic DNA is of at least as much interest from this perspective as true functional genes. See appendix for a brief guide to some of the genetic terms to be used in this paper.

3 In this paper I use ‘genetic’ and ‘genetics’ in the biological sense. To avoid confusion, I will follow Ross (2005) in using ‘genealogical’ where many linguists might use ‘genetic’.

59
A classic study in this field is that of Serjeantson et al. (1983), which sounds a note of caution. Using multivariate analysis of classical autosomal (non-sex-chromosome) polymorphisms as their measure of genetic variation, and shared cognates (a lexicostatistical measure not based on shared innovations) to quantify linguistic relationships, they examined linguistic and genetic differentiation amongst seventeen language groups (both Austronesian and Papuan: see below) in Bogia sub-province, north coastal New Guinea. They found that the genetic and linguistic measures of difference were strongly correlated, and both were strongly negatively correlated with geographical distance, all as expected. But when they used partial correlation to control statistically for the effect of one variable while examining the relationship between the other two, they found that the gene-distance relationship was still quite strong when the effect of cognates was held constant, but the gene-cognate relationship became negligible when distance was held constant. They concluded that, at least in this case, the apparent gene-cognate relationship was secondary to the dependence of both genes and cognates on geographic distance.

Serjeantson et al. also reviewed earlier literature as well as their own findings on the overall contrast between speakers of Austronesian and Papuan languages (i.e. languages in either the Austronesian family or one of the numerous families in the region grouped together by negative definition as non-Austronesian or Papuan). Earlier research had identified genetic differences between speakers of the two language groupings in the upper Markham Valley, New Guinea. The same differences, however, were not reproduced when Austronesian- and Papuan-speakers in Island Melanesia were compared. The authors drew attention to the possibility that the differences observed in the Markham Valley might have been primarily highland/lowland differences. More generally they argued that the genetic data in their study do not strongly signal the presence of a major linguistic difference, and that to some extent populations now speaking Austronesian languages may have substantial biological ancestry from Papuan-speakers.

A microcosm of the relationship between Austronesian- and Papuan-speakers is seen between two of the language groups included in Serjeantson et al.’s study, the Austronesian-speaking Takia and Papuan-speaking Waskia, both living on the small offshore island of Karkar. Here, as Boyce et al. (1978:292) concisely express the outcome of their genetic distance analysis, ‘there is no overlap between the clusters of Waskia and Takia village groups, but neither is there any marked separation between them.’

In the following sections I shall return in more detail to the Papuan and Austronesian languages of the New Guinea area, and their speakers, but first let us range more widely in pursuit of examples of parallelisms or otherwise between genetic and linguistic differentiation: Europe suggests itself.

A series of major papers has addressed the question of genetic differences between linguistically defined populations in Europe. Sokal and colleagues (Sokal 1988; Sokal et al. 1992) similarly analysed the inter-relationships of genetics, language and geographical distance; and found speakers of different language families to differ genetically as in New Guinea, but in this case even when geographic differentiation is allowed for. His group also found greater biological difference, more marked for gene frequencies than for cranial variables, across language-family boundaries than across similar distances elsewhere (Sokal et al. 1988). Barbujani and Sokal (1990) confirmed that zones of sharp genetic change in Europe tend to coincide with language family, language or dialect boundaries. They argued that language affiliation, more than environmental factors, helps maintain and probably cause genetic differences. Sokal et al. (1989) confirmed that European
populations differ more among language families than within them. Rosser et al. (2000), however, analysing Y chromosome genetic markers (inherited in the male line), found results pointing to the primacy of geographical factors in shaping genetic variation, though linguistic factors were accorded some role.

Harding and Sokal (1988) computed multivariate genetic distances which reflected geographical proximity (hence short-range gene flow) more strongly than linguistic relationships, except that where linguistic contrasts were at a higher level — Basque-, Finnic- and Semitic-speaking groups compared with their Indo-European-speaking neighbours — genetic relationships were more distant. Calafell and Bertranpetit (1994) focused on the Basques and interpreted their multivariate analysis results as consistent with archaeological and linguistic indications of long in situ differentiation, perhaps since the last Glacial Maximum. At a more micro level, linguistic (or dialectal) differences within Italy seemed to show a better correspondence with zones of sharp gene frequency change, i.e. they were apparently stronger barriers to gene flow, than did geographical barriers (Barbujani and Sokal 1991a, b; Barbujani et al. 1992), though this was not the case with mitochondrial DNA (inherited in the female line) (Barbujani et al. 1996). Manni and Barrai (2000) have confirmed the relationship between genetic and linguistic boundaries in Italy, even in a small-scale case study.

Political borders, too, and historical events as recent as the Second World War, have been found to correspond to borders observed in the genetic variation patterns (Kayser et al. 2005).

These findings taken together present a complex picture; but it is a reasonably encouraging one in suggesting that there can be an isomorphism between genetic and linguistic patterns of differentiation, sufficiently often to justify continuing investigation into the possibility elsewhere than Europe, including, of course, New Guinea.

Linguistic variety in New Guinea

So often has attention been drawn to the unparalleled linguistic diversity of the New Guinea region that it has become a cliché. Yet, with some 1000 languages spoken on and in the region of an island that is not much larger than Pakistan in its land area (and much less populous), the phenomenon is genuine (Foley 2000). And the challenge it poses for linguists, to discern the patterning of this diversity and offer an interpretation in historical terms, is correspondingly extreme — the more so for the potentially very deep antiquity of its origins.

Of the 900 or more languages actually spoken on New Guinea and the small islands immediately offshore, some 150, mainly coastally distributed, belong to the Austronesian language family mentioned above (Pawley 1998). This family is well demarcated and reached the region relatively recently (~3500 years ago). It is the relationships amongst the 750 or so Papuan languages that represent probably New Guinea’s greatest challenge to historical linguistics. These languages are not only numerous relative to the size of the territory their speakers occupy: they are also markedly different from one another. Very seldom has anyone considered them all genealogically connectable into a single language family. Ross (2005) groups them provisionally as 23 distinct unconnectable families (in roughly the same sense that Austronesian or Indo-European are families), plus nine or so isolated languages that cannot be grouped with any other(s) into families. In his family-building, Ross is neither as bold as Wurm (1975) was, nor as sceptical as Foley (1986, 1992).
The Trans New Guinea family

Amongst attempts to discern patterns amidst this bewildering variety, McElhanon and Voorhoeve’s (1970) proposal of a Trans New Guinea (TNG) language phylum has proven both productive and controversial. Under this proposal, a large number of Papuan languages, covering a number of the previously recognised families and ranging across much of the island from southern West New Guinea through the central highland cordillera to Madang, Morobe and even Northern Provinces of Papua New Guinea, can be connected together, albeit more tenuously, into a larger and higher-level genealogical grouping, with an implied unitary ancestor language. This grouping is the hypothesised TNG phylum or family. It has been linguists’ acceptance or otherwise of this proposed family as a genealogically coherent grouping which has most radically affected their enumerations of families within the overall Papuan category. Assessments accepting the TNG grouping as valid have been able to combine a number of the sceptics’ families into it and so to list fewer Papuan families overall.

Prominent amongst the scholars to take up the challenge of evaluating and refining the TNG proposal has been the dedicatee of this festschrift, Andrew Pawley. He describes (1998, 2005) how the scope of the TNG hypothesis was greatly expanded by its originators and others, from an initial 130 languages (TNG I) to a set of 256 (TNG II) regarded by 1975 as belonging securely within the TNG grouping, and less confidently to a larger set still, numbering 491 languages in total (TNG III). He also outlines a number of quite trenchant criticisms made of the TNG hypothesis in the later 1970s and 1980s, on technical linguistic grounds such as the proponents’ failure to implement the comparative method rigorously. This left the proposal in a limbo, unproven and to some extent neglected amongst linguists, even around the time when it was incorporated in its most ambitious form into Wurm and Hattori’s (1981–83) spectacular and widely known Pacific language atlas.

But since the mid 1990s, as Pawley also shows in recounting the ‘chequered career’ of the TNG hypothesis, a reassessment has been under way. This was partly at his own instigation: by then he had had cause to reconsider his initial acceptance of the prevailing critique. Meticulous comparative analysis of the data since that time has persuaded many specialists that the TNG hypothesis is essentially correct — even if the methodology and arguments offered in its favour in the 1970s were not good enough to convince; and even if the version now reinstated is not exactly the same as any of those set out at that time. Pawley (2005) refers to the reinstated version as Trans New Guinea IV. Not only he and Ross (2005) but also Foley in his more recent work (2000) are now satisfied — albeit with varying emphases and issues yet to resolve — that a TNG family, broadly TNG IV, can be justified. In numerical terms, its coverage is greater than the TNG II of Wurm and his colleagues, but not as extensive as their TNG III. On Ross’ (2005) reckoning, at least 311 of the 605 Papuan languages for which he had adequate data, and possibly up to 359 of them, can, with varying methodology and confidence, now be assigned to the TNG family.

Foley (2000), who put the number of TNG languages at close to 300, estimated that around two million people, roughly half the total Papuan-speaking population at the time, spoke TNG languages. And he characterised the family as primarily montane in its distribution, though spilling out from the mountain valleys into the highlands fringes and sometimes the lowlands too, most extensively the southern lowlands. Thus the great majority of non-TNG Papuan language families is found in the northern half of New Guinea; and, whether reckoned in languages, speakers or territory inhabited, all these families are
substantially outnumbered by the TNG family — as is New Guinea’s second largest language family, Austronesian. The numerical preponderance of the TNG family seems only likely to become more striking as the work of historical linguistics proceeds and more languages and families currently unexamined or in doubt are shown to belong to it.

How does a language family become widely distributed?

In the growing consensus that we now have on the indigenous New Guinea linguistic scene, then, the predominance of the TNG family emerges as an exceptional and salient feature of the pattern. TNG may not have attained the extensive global geographical distribution of the Austronesian family (whose New Guinea members form only a small fraction of a total distributed from Madagascar to Hawai’i and Rapanui/Easter Island), or for that matter of the Indo-European, Bantu, or Sino-Tibetan families. But in the New Guinea context, where other Papuan language families comprise only 25 languages on average and most languages have fewer than 3000 speakers spread over 10–20 villages (Foley 2000), the dispersion of TNG is remarkably wide. Foley offers some suggestions as to why other Papuan language families are quite restricted, spatially and demographically — the degree of differentiation permitted by great time depth, the politics of small-scale social structures, the ‘residual’ character of the New Guinea region, esoterogeny. But the point I wish to pursue here is how TNG became the exception that it has increasingly clearly been shown to be.

Some fundamental and fascinating questions arise about cultural and population history, and Pawley (1998, 2005) amongst others has posed a number of them. Thus for example:

- What cultural factors enable TNG languages to attain such a wide distribution compared with other Papuan families?
- To what extent were the languages carried along by colonising populations as opposed to spreading among already established populations?
- Where were the initial dispersal centre and directions of spread? When did the dispersal(s) occur? … Was knowledge of agriculture the cultural advantage that allowed TNG speakers to spread over a large part of New Guinea? (1998:683–684).

In this contribution, I as a non-linguist must leave any remaining questions about the characterisation of the TNG language family to the specialists. Rather, I simply accept the consensus view as just set out. My aim is instead to think aloud about the cultural, demographic and genetic implications of this view, in the light of Pawley’s questions above, especially the last one quoted, later expressed as: ‘Was the TNG expansion powered by agriculture?’ (2005:97). In doing so, I shall look to other cases elsewhere in the world, especially Europe again, as well as New Guinea itself. I should be clear before I proceed any further that I shall not be able to answer this question from the genetic perspective that will be my main focus: the necessary data are simply not available yet. I shall — necessarily — allow myself to be speculative, and shall proceed by asking a series of distinct but linked questions.

As to the date of the TNG dispersal, Pawley (1998, 2005) and Foley (2000) are agreed — though neither would wish to see too much credence placed on glottochronological methods — that the strong internal diversity of the family, structurally and lexically, implies deep antiquity. Foley saw TNG as considerably more ancient than the Oceanic subgroup of Austronesian, estimated at ~3500 years old. Pawley (1998) suggested a dispersal at least 5000 years ago and possibly earlier, and went on (2005) to point out that the Madang subgroup alone of TNG is more diverse lexically than the whole of either Austronesian or Indo-European, conservatively estimated to have broken up over 4000 and over 5000 years ago respectively. While pointing out that a literal glottochronology might put the break-up of
the TNG protolanguage as early as 10,000 years ago, Pawley restricts himself to the more cautious claim that TNG is probably older than Austronesian or Indo-European. Ross (2005) considers a date of 6000 years ago plausible for the dispersal of TNG.

And as to the dispersal centre, the main appeal is to the argument that this can most plausibly be located in the area of greatest linguistic diversity today, viewed especially in terms of the density of high-level subgroups. For Pawley (2005), this logic identifies the region of Papua New Guinea from the Strickland River in the west, eastwards to the Eastern Highlands and Madang Provinces plus the Finisterre Ranges and Huon Peninsula, as at least an early dispersal centre — not necessarily the very start of the process. Foley (2000) is more specific, identifying the Eastern Highlands or Upper Markham valley as the probable homeland of the TNG family, in turn making it likely that the main direction of subsequent dispersal was westwards, through the central cordillera of Papua New Guinea and West Papua, into the southern lowlands, and eventually reaching the eastern Indonesian islands of Timor, Alor and Pantar.

Agriculture in New Guinea is considered an indigenous development, evidenced with some definiteness from around 6500–7000 years ago in the highland valleys east of the Strickland (Golson and Gardner 1990; Denham et al. 2003; Denham 2005). Thus both the timing and the geographical origin of the TNG dispersal appear, in so far as present evidence allows any specific statements, compatible with Pawley’s linkage of that dispersal with early agriculture. It is that link between language family and agriculture, and the association of both with actual population expansion, that I wish to consider in greater detail here.

From a linguistic perspective, Pawley (1998) — like Bellwood (1978, 1997) from an archaeological perspective — has suggested that the dispersal of the TNG languages (and possibly the populations speaking them) may have been powered by an agricultural expansion. Analogous suggestions regarding other parts of the world have been in the air for a long time. In particular, the literature on the European case is much more extensive and I shall turn briefly to it here, with a view to what lessons it may offer for the New Guinea case. Indirectly, Ammerman and Cavalli-Sforza (1984) got the ball rolling, with their argument that the diffusion of early agriculture into and across Europe was not only cultural but also demic (involving actual population movement), and involved a ripple of change which took over two thousand years to expand west, north-west and north from the south-east. The process is suggested to have involved, not heroic treks, but a frontier expanding at an average rate (not precluding wide variation) of 25 kilometres per generation.

Ammerman and Cavalli-Sforza (1984) envisaged the transition from hunting and gathering to farming as involving a complex of social processes taking place at the leading edge of the ripple. These, they propose, would have included: local migration of farmers into non-farming areas; acculturation of indigenous local hunter-gatherers to agricultural ways of life; intermarriage; conflict and warfare; spread of infectious disease; and mutually beneficial interactions of various kinds. That the net outcome of these encounters was eventually that all Europeans became farmers — rather than the invaders retreating, dying out, or becoming hunter-gatherers — depended crucially on significant population growth at the edge of the ripple, permitted by the higher population carrying capacity characteristic of agricultural rather than hunting/gathering systems. Ammerman and Cavalli-Sforza inferred, plausibly, that if their model were correct, the pattern of human genetic variation across Europe at the end of this process should take the form of an arc-shaped cline
(gradual spatial change in gene frequencies) with a southeast-northwest orientation. The strongest single pattern to emerge from their multivariate analysis of twentieth century genetic variation amongst Europeans (the first principal component, accounting for 27 percent of overall variation) was — despite everything that has happened in European history in the intervening millennia — a cline conforming quite well to this expectation. On this basis, they declared their model supported by the findings.

Ammerman and Cavalli-Sforza’s argument encountered both support and scepticism, as well as elaboration when new genetic methods and data became available (see below). But they were near-silent on linguistic questions—only passingly raising the possibility that a subsidiary pattern in their analysis (the third principal component, accounting for 11 percent of the variation) might reflect a quite separate diffusion process, a post-neolithic expansion of Indo-European (IE) speakers from an area north of the Black Sea, consistent with the prevailing consensus among Indo-Europeans, then and to some extent still. It was the archaeologist Renfrew (1987, 1988) who subsequently suggested that Ammerman and Cavalli-Sforza’s demic diffusion of agriculture might actually have been the vehicle for the European diffusion of the IE language family as well.

For Indo-Europeans, Renfrew’s suggestion was important in challenging the orthodoxies regarding both the date and the geographical origin of the IE language family. Debates sprang up as to its plausibility on various historical linguistic grounds including the reconstructed vocabulary of Proto-IE. The antiquity, provenance and specific historical diffusion routes of the IE language family are, however, not my prime concern here. What I wish to pursue, in the context of the analogy with TNG languages, is a different aspect of Renfrew’s hypothesis — the fact that, as Pawley does for the TNG family, he posits a specific driving force for that diffusion.

Why agriculture?

If we follow the Ammerman/Cavalli-Sforza model, the demography is crucial, even if in itself it has attracted less discussion than the agriculture, the linguistics or the genetics. The demography of extant hunter-gatherers (e.g. Peterson and Taylor 1998; Pennington 2001; Gurven and Kaplan 2007) is not reliable as a guide to that of hunter-gatherers who lived in the remote past and on different continents, of whom they may be in several ways unrepresentative; but it is the only direct information we have. Much the same might be said of the demography of small-scale agricultural societies, even though there are more of these still extant. But with such grounds for caution acknowledged, it still seems reasonable to work on the basis that in most situations hunter-gatherers have lived at lower population density than their immediate successors. Even if estimates that put the differential at up to 50-fold are exaggerations, relative to the conditions prevailing in the European neolithic transition, the margin for error is considerable.

Whether the introduction of agricultural materials and practices simply triggered a rapid population growth from the hunter-gatherer carrying capacity to the higher agricultural one, as Ammerman and Cavalli-Sforza seem to assume, or conversely population pressures also played some part in driving adoption of agriculture, is perhaps not crucial to the linguistic or genetic outcomes. What matters in the present context is the association of a phase of population growth with early agriculture.

The implication is that, despite the pre-existing hunter-gatherer population, there was effectively a vacant or at least under-filled niche open to the first farmers, at the leading edge of agriculture’s diffusion. Here, conditions may have been favourable enough for
their fertility to exceed their mortality substantially for a time. It may be easier to imagine factors favouring increased fertility than reduced mortality, though the effects on family formation from reduced infant and child mortality or raised fertility would be comparable. Perhaps changes of both kinds played a part. Convincing parallels for this are hard to find, although Bellwood (2005) has drawn attention to the example of transiently very high fertility of colonial populations in nineteenth-century Australia (see also Bocquet-Appel 2009 and sources cited therein). A plausible (though doubtless not uncontested) model might propose an abundance of adequate (if less varied) food, reduced inter-birth intervals related to the greater availability of weaning foods in particular, and a greater variety of child care options permitted by a more sedentary lifestyle — perhaps also an ability to exploit environment types not so well suited to hunting and gathering. While any difference in military technology seems unlikely to have been decisive in determining the outcome of hunter-gather/farmer conflicts, a numerical advantage once developed would assist farmers in settling new territory whatever the wishes of the indigenous people, and in defending food and other resources against them. This would in turn fuel further demographic and territorial inequality between the two groups for as long as they continued to have separately recognised existences.

Crucial to the Ammerman/Cavalli-Sforza model too is that interactions between the groups need not only have been hostile. Other kinds of interaction may have included exchange systems covering material goods and local knowledge, and intermarriage, as well as the acculturation already mentioned. Indigenous spouses who married into early farming communities, and perhaps too whole indigenous communities who adopted agriculture and became part of regional social structures, would have played the key role of introducing, at each stage of the process, indigenous gene flow into what became the farming population. This is what differentiates the model from both of the extreme alternatives, on the one hand where diffusion is purely cultural without population movement, and on the other where the indigenous population is comprehensively put to the sword. And it is the repetition of the same process over perhaps 65–80 generations, around the agricultural frontier as it expands across Europe, that gives the genetic outcome its clinal character. In each place and each generation, the majority farming population is of mixed descent, the proportions of the mixture changing gradually as the ripple of agriculture spreads, in such a way that in the original farmers’ genes are more and more diluted by indigenous European genes as the process progresses.

As Ammerman and Cavalli-Sforza note in passing and Fix (1999) discusses in detail, an agricultural diffusion process might also generate a cline in a different way — by any natural selection imposed by novel features of the agricultural way of life itself, for example changes in diet, a process which would have been at work for longer and so may have gone further in the south-east than the north-west. But this would only affect certain genes and may be a side issue here.

And what then of the linguistic changes that might have accompanied the processes which Ammerman and Cavalli-Sforza model? These would be IE languages, if we allow with Renfrew that early farmers in Anatolia spoke early IE languages, and were their vectors into Greece and thence into the Balkans, Italy and much of the rest of Europe. What would happen to the language or languages of the invading farmers (IE or not) under the circumstances envisaged by Ammerman and Cavalli-Sforza, as their subsistence practices and materials and, with some dilution, their genes diffused across a Europe generally agreed to have spoken non-IE (possibly Basque-related) languages until that time?
We can infer very little directly about the interaction of speech communities in the remote past, even if we accept the premise that the context was as outlined by the model discussed here. It does seem plausible, however, that, where interaction is local, small-scale and face-to-face, languages spoken by a majority may be the likeliest to survive in the long term. And if we can accept the argument so far, it will usually have been the farmers that were in the majority demographically. Likewise, if they were in the majority, they will, ceteris paribus, have left more descendants and a bigger impact on the gene pools and languages of later generations.

Thus what gives the Ammerman/Cavalli-Sforza demic diffusion model of agriculture its linguistic and genetic explanatory power is its demographic base. This makes it plausible that it is the actual movement of speakers into new areas in significant numbers that allows a language family to become widely distributed, along with the genes of its speakers. All language families presumably originate with an ancestor language spoken in a relatively small, spatially and numerically restricted population. It is the coincidence of their presence in a population undergoing a numerical and territorial expansion at the time that allows a language and its descendants to become widespread geographically. And, following the geographical expansion of a population speaking the ancestor language, or languages immediately descended from it, the conditions are right for break-up and differentiation into a larger number of eventually more distantly related languages, with a collective dispersion potentially continental in scale.

In keeping with this scenario, Bellwood (2001, 2002, 2005; Diamond and Bellwood 2003) has strongly argued that, as a global pattern, independent origins of agriculture have repeatedly led to the growth and territorial expansion of early agricultural populations — an expansion which took with it not only people and their agriculture but many other cultural elements, including material culture and language, and also the genetic markers that happened to characterise each population affected. Most often (not invariably) the converse also applies: that where a language family has attained a widespread geographical distribution, it is agriculture which has driven this process, at least where the expansion is into territory already occupied by other groups. Amongst the exceptions which appear to challenge this latter proposition is the Pama-Nyungan family of Australia. Bellwood’s main concern, however, is to contrast his model with the notion that agriculture diffused through pre-existing populations without movement of people. Bellwood is also clear that he sees these as complex expansions which biologically incorporated the surrounding hunter-gatherers rather than totally obliterating them, while profoundly changing their cultures in many respects.

Bellwood views this type of outcome as the general pattern to be seen across a broad range of cases of agricultural and language family origins, a generality not contradicted by the fact of variation amongst them. Southwest Asia, China, Mesoamerica and Sub-Saharan Africa supply his main case studies. He devotes considerable space to the European case, as well as its ultimate Southwest Asian source. Although he sees the involvement of IE dispersal in the diffusion of agriculture across Europe as still a vexed issue, ultimately he inclines in favour of it (Bellwood 2005), citing amongst other research the computer simulation of Barbujani et al. (1995) and the quasi-evolutionary linguistic analysis of Gray and Atkinson (2003). New Guinea is not one of Bellwood’s main case studies and his comments on it are brief. He notes that the New Guinea highlands’ transition to agriculture, while an autonomous local development of great antiquity, did not have an expansive character on the scale that he describes for several other parts of the world —
meaning that there is no evidence for its having driven a dispersal to, for example, Island Melanesia or Australia. He speculates that rugged terrain and, in more densely populated parts of the lowlands, the toll taken by malaria may have applied a brake to any expansive tendencies.

As discussed earlier, however, the TNG language family clearly became widespread to an extent exceptional within New Guinea; and the societies in question generally practised an agriculture not obviously traceable to any origin external to it. Ultimately, my interest here is in whether an early agricultural population expansion within New Guinea may have left traces in both the languages and the genomes of present-day New Guineans. At the present time, since the European case is much more thoroughly researched, the debates about Europe may point the way to develop any parallel there may be.

**Does the model work for Europe?**

The previous section was predicated on acceptance of the Ammerman/Cavalli-Sforza model. In it I attempted to spell out some implications in terms of the social process, so that readers may imagine for themselves how applicable they might be in New Guinea. But should we in fact accept the model? Does it work, from a genetic point of view, even for Europe? What level of confirmation can even fulfilment of genetic predictions provide for the archaeological and linguistic facets of it?

To address such questions even if only briefly, let me rehearse some of the reception and follow-up research, with particular attention to the genetic and other human biological research. In doing so, I shall take it for granted that some simplifications in the model are conscious and intended as heuristic, rather than just erroneously mechanical: for example, that the stated rate of the wave of advance, 25 kilometres per generation, is to be read as an average and not as an assertion that diffusion proceeded at a constant rate regardless of the landforms and other environmental and human factors encountered. I shall also take it for granted that no-one would expect all of the genetic variance observed to be explained by a single ancient migration process. If traces of such a process are discernible at all when so many other, mostly more recent, historical factors have also been at work, that is striking enough: it is not to be expected that one process should explain the appearance of the whole geographically complex palimpsest.

The research group of Sokal followed up the original modelling and multivariate analysis with a series of studies, summarised by Sokal (1991), which took the ideas further, using somewhat different techniques, still applied to classical nuclear autosomal genetic polymorphisms. They were able to confirm the existence of clines similar to those of Ammerman and Cavalli-Sforza, and additionally to demonstrate that these had statistical significance. They were also able to dissect the spatial patterns gene by gene, which yielded complex results. Some genes showed no significant pattern, and not all of the significant patterns were clear clinal trends; but the majority of the clear clinal trends (9/16) had the northwest-southeast orientation predicted by the model. This was generally construed as confirmatory. Indeed, when migrants and indigenes may not have differed in some gene distributions before the diffusion started, or may have differed but in complex ways, it seems reasonable that a signal from the postulated diffusion would not be equally clear on all variables examined. They did not find similar support for the model in craniometric measurements, though they did not find disproof either (Harding et al. 1989).

When the DNA of the human mitochondrial genome became well enough known to become a research tool in anthropological genetics, and was applied to the European
Neolithic problem especially by Richards and others in the Sykes research group, results initially seemed contradictory to the model. Clines were not detected, many lineages were inferred to be pre-neolithic, and increasing doubt was cast on whether there was a significant demic component to the diffusion of agriculture. More recent mitochondrial DNA research with larger sample sizes has, however, shown evidence of clines in this marker system also, though the interpretation of it is not straightforward (Richards et al. 2002). Analysis of variation in the Y chromosome has yielded results containing what is interpreted as a strong signal of expansion from the Near East, though complicated by other factors also (Rosser et al. 2000). Richards (2003) summarises both recent mitochondrial and Y chromosome research as suggesting that both indicate a neolithic Near East contribution to the European gene pool of around one quarter or less. He views this as demographically minor: but maybe supporters of the model would be quite satisfied to have demonstrated such a contribution. Much of the difference between viewpoints may lie in the wording and emphasis. A number of authors see merit yet in some form of the demic diffusion model (e.g. Barbujani and Bertorelle 2001; Barbujani and Dupanloup 2002; Chikhi 2002). The model is, then, still under debate; but it remains in play. In the present context, it still seems reasonable to accept it as a useful reference point for purposes of stimulating ideas about New Guinea.

Could any other process have similar effects?

It seems plausible that agriculture might under the right circumstances diffuse either demically or culturally. Given that varying subsistence systems have varying material consequences, there could be tangible, context-dependent advantages to adopting the practices and materials of agriculture, whatever the means by which they become available to the people making those choices.

Under what circumstances, though, might a language or language family diffuse, if not demically? I am no more a sociolinguist than a historical linguist and the following discussion necessarily reflects that. To anyone primed to think in terms of material consequences, the advantage to switching one’s linguistic affiliation seems harder to imagine. Why would people do that, if not under pressure of some kind? It is not that language shifts never occur without major population flow. They are documented, for example, amongst African and Asian hunter-gatherers living adjacent to agriculturalists, as well as in large parts of Africa and Asia under modern conditions. But in general these are cases involving some kind of inequality — be it economic, political, technological or military — between the longer-standing speakers of the new language and its adopters. Inequalities based on unlike economic systems, or possession of for example horses, or guns, or sheer political inequality, as where ‘élite dominance’ is exerted in a stratified society or nation state leading to the suppression of Celtic languages in historical Europe for example, might certainly induce language shift in a less powerful group.

But most of these inequalities, other than agriculture, are difficult to envisage acting on a substantial scale for the New Guinea case. Again, I do not wish to overstate the case. Where local linguistic diversity and bilingualism are high, there may be shifts of linguistic affiliation at a village level within a generation or two; and languages may be prized and adopted for reasons connected with prestige, exchange and identity (Foley 1986). And where genetic heritage indicates one affiliation and linguistic heritage another, such as on the New Guinea island of Mailu, there may have been a linguistic change, or a linguistic affiliation may be retained whilst the society receives a substantial in-migration from a
genetically contrasting one, perhaps a regular exchange partner (Kirk 1992). A long-term trickle of migration might be imagined leading to the receiving population retaining its principal language, since its speakers are in the majority at any one time, but in the long run being significantly influenced genetically by the donor population. Again, avoidance speech styles as traditionally sound in Australia may encourage linguistic borrowing from neighbours. But could any of these processes so favour one language or language group over others as to lead to its dispersion across two thousand kilometres?

As Bellwood argues, issues of scale are important. It seems doubtful whether social processes of these kinds, at the level of individuals, households or villages, in small scale societies not inherently or permanently subjugated to one another, can really explain language shifts on a continental or large-island scale, or the expansion of single language families, such as Indo-European, Niger-Congo, Sino-Tibetan, Austronesian or perhaps TNG over very great distances. There is much in the historical synthesis of Ostler (2005) to support Bellwood’s position in this regard, that major language family expansions tend to be fuelled by significant movement of their speakers, even though not to the exclusion of others. Further, while ‘élite dominance’ can disseminate a language or group of languages widely for a while, it usually declines when the élite declines, unless sufficient speakers have by then joined the general population. All this seems to bolster the plausibility of the demic diffusion argument as a candidate to explain the geographical reach of the TNG family amongst others.

Agriculture, TNG languages and the genetics of their speakers

A tenable working hypothesis then is that the population dynamics characteristic of agricultural systems drove a numerical and territorial expansion millennia ago of certain New Guinea (probably highlands) societies, together with the particular (TNG) languages which they happened to speak and the genetic markers that happened to be distinctive of them; and that diversification, both linguistic and genetic, then followed.

This is a complicated proposition conceptually, and testing it is not straightforward. Can we take agriculture’s different manifestations worldwide to represent a sufficiently unitary phenomenon that we can apply demographic expectations formed in one part of the world to a very different one? Or are there important properties, material or cultural, to New Guinea’s tuber-based agricultural system which might make it importantly different in population dynamics or other respects relevant in this context? If we do detect genetic signals of population expansion, what would be required to associate that expansion with agriculture and/or the TNG expansion? If we assume that TNG-speaking agriculturalists expanded into territory already occupied by hunter-gatherers, was the interaction of the complex type envisaged by Ammerman and Cavalli-Sforza, resulting in each place in a population composed of both locals and incomers?

Presumably populations not yet reached by the TNG expansion at a given point in the process were at that time hunter-gatherers, so do we need to develop a sub-hypothesis to account for the widespread occurrence of agriculture amongst non-TNG groups by recent times? Did they develop or acquire agriculture independently of the TNG groups, or did the TNG/agriculture expansion reach most of the island, somehow losing its TNG component at some point along the way? One argument might be that it is precisely the (mainly lowland) regions beyond the distribution of the TNG languages, for example in the Sepik basin, which are most often home to New Guinea’s documented hunter-gatherers and other groups making greatest use of wild foods (Roscoe 2005); and that these regions
may also be where agricultural practices have most recently been acquired, putatively by cultural diffusion from TNG-speaking neighbours in the first instance, and where even now commitment to an agricultural way of life is least. All of these questions, and others, raise problems for the debates on this hypothesis.

But the most basic difficulty, and the one which on its own precludes any firm conclusion here, is the sheer paucity of relevant data, especially on the genetics. By way of contrast, Harding and Sokal (1988) were able to include data from 3369 local populations in their study of genetic distances amongst speakers of five European language families (phyla, in their usage). Disappointingly often (e.g. Redd and Stoneking 1999; Hujdajov et al. 2007), New Guinea, the paragon of linguistic diversity, is represented in anthropological genetic studies by only one or two (‘highlands’ and ‘coast’) local populations, whether linguistically defined or not, or a pooled overall one.4

Amongst studies which sample New Guinea most thoroughly are those of Tommaseo-Ponzetta et al. (2002) and Kayser et al. (2003, 2006) for West New Guinea (WNG), and Harley (Harley et al. 2005; Harley 2006) for mainland Papua New Guinea (PNG), i.e. eastern New Guinea. Between them they cover a selection of markers in both mitochondrial DNA and the Y chromosome, as well as some autosomal DNA markers in the case of Harley.

None of these studies was expressly designed to test our central hypothesis here, nor to detect genetic signals of past population growth, for which there are several specific methods (Excoffier and Schneider 2000; Jobling et al. 2004), not deployed in these studies; but all took trouble to identify the linguistic affiliation of participants at each study location. As a first step, it might seem logical to set up a prediction that if the TNG language family has been dispersed by demic diffusion along with New Guinea agricultural systems, then TNG speakers now would be more similar genetically on average to each other than to non-TNG speakers, at least once any confounding with geographic distance between populations has been controlled.

Unfortunately, as a by-product perhaps of the TNG family’s sheer size and territorial coverage, Tommaseo-Ponzetta et al. (2002) and Kayser et al. (2003) only surveyed TNG-speaking populations for their studies, and so no comparison with non-TNG-speaking populations can be made. Kayser et al. (2003:299) find that ‘in general, there does not seem to be a significant correspondence between language and the Y chromosome in WNG’; but this conclusion is clearly relative to its purely within-TNG frame of reference.

Harley’s (2006) sampling design has the best potential to examine correlations between genetic and linguistic variables, even though it was not so intended either. It covers 32 village populations selected from across mainland PNG, corresponding to nineteen language groups, of which thirteen fall into the TNG family, four into Sepik-Ramu, one into Torricelli and one for comparison into Austronesian. Harley found that, while mitochondrial DNA variants displayed a great deal of variety mostly occurring at the within-village level, with very little regional or linguistic signal in the variation patterns, the opposite was the case for the Y chromosome, in that the observed variation was markedly geographic and any sampled male ‘could be placed within his village with 80

4 Of course many of these studies are designed at a macro level, to compare New Guinea with other regions elsewhere in the world, and not to address the relatively micro questions addressed here. Nonetheless, the difficulty in matching the abundance of genetic sampling attained in the best studied regions, or available information on linguistic diversity in New Guinea, to genetic sampling in New Guinea constitutes an obstacle for the field at the present time.
percent accuracy, language group with 99 percent accuracy, or language family with 100 percent accuracy’ (p.58). In this context ‘family’ refers not to TNG but to its component parts such as Ok and Chimbu-Wahgi and to other groupings at similar level such as Sepik Hill.

Harley focuses mainly on the gender-related interpretation of this contrast. But for the present purpose, the salient point is that mitochondrial DNA supplies a very weak measure of differentiation, and the Y chromosome in a sense too strong a one: the differences between language groups are so absolute that it becomes difficult to say meaningfully that one difference is greater than another. In Harley’s main Y chromosome analysis, a large majority of all pairwise village comparisons show a highly significant difference (Tables 4.6, 4.9, 4.12). Amongst the few which do not, village pairs sharing the same language are greatly over-represented as one might expect. But overall the differentiation is so strong that it is hard to draw inferences from this data-set about relative genetic inter-relatedness at higher level of linguistic inter-relatedness.

In Harley’s results, the autosomal (bi-parental) genetic markers, like the mitochondrial ones, contain a much weaker regional/linguistic signal than the Y. But this paradoxically opens the possibility of greater informativeness about higher-level relationships: of all the pairwise comparisons between the 19 language groups, 15/78 (19 percent) intra-TNG comparisons yielded non-significant results; 9/78 (12 percent TNG/non-TNG) comparisons yielded such results; and (strikingly) none of the non-TNG/non-TNG comparisons did so (Table 5.6). An optimist might see in such results enough consistency with an agricultural demic diffusion model to be encouraged to take the idea further, though a control for geography remains to be applied. For the time being, Harley’s realistic caution remains in order and she deserves the last word: ‘… linguistic/geographic hierarchical definitions for Papuan-speaking peoples have some credibility, in terms of being a biologically meaningful population definition, at the village and language population level. Beyond these levels, i.e. at the language family level, however, the populations defined in terms of groupings of speakers have proven hard to characterize in genetic terms’ (p.162).

In sum, then, the central idea drawn from Pawley (1998, 2005) for review in this chapter remains a fascinating and stimulating one; but at this stage it is ahead of its time in terms of what the genetic evidence can yet corroborate or disprove.

Appendix: some terms and concepts in anthropological genetics

The complex and technical nature of modern genetics makes for recognised communication obstacles in cross-disciplinary contexts like the present one. It is not entirely possible to overcome these obstacles in an appendix such as this, nor even in a longer piece such as provided by Attenborough (2005). Nonetheless, the following notes are offered in the hope that they may be helpful.

*Gene* is a term variously defined since before there was any understanding of the molecular basis of genetics. Essentially it now refers to a sequence of deoxyribonucleic acid (DNA) which codes for the synthesis of a protein that plays a part in body function — a different sequence for each different protein. The 23 pairs of *chromosomes* found in our cell nuclei contain our 32,000 or so different genes, but our chromosomes contain much more DNA than that would require. Most of our DNA is not part of any structural gene as just characterised, that is it is *non-genic* (or non-coding). Some non-genic DNA plays a functional role of another kind, for example in regulating the level of gene activity; but a large fraction of it has no currently known function and may have no function at all.
Of our 23 chromosome pairs, 22 are known as autosomes, and the DNA they contain is said to be **autosomal**. It is this DNA, the great majority of our DNA, to which Mendel’s principles, as modified by subsequent knowledge, primarily apply.

The 23rd pair — known as the sex chromosomes — is of special interest on account of its role in biological sex determination. In general, individuals whose two sex chromosomes are both of the type known as X are biologically female, while individuals who have one X chromosome and one of the smaller type known as Y are biologically male. **Y chromosomes** are thus transmitted in patrilineal mode from father to son only.

Fortunately there is an approximate matrilineral counterpart to the Y chromosome, in the form of the tiny fraction of human DNA which occurs not in cell nuclei at all but elsewhere in the cell, in bodies called mitochondria and thus known as **mitochondrial DNA** (or mtDNA). Mitochondrial DNA, though present in both sexes, is transmitted only by females, leading to the formal analogy with Y-chromosomal inheritance.

These non-Mendelian forms of inheritance, both mitochondrial and Y-chromosomal, allow geneticists to deploy powerful analytical approaches based on the tree-like patterns of connexion between single-sex ancestors and their descendants in each case. Where analyses of both forms show the same populations in contrasting lights, this may suggest interesting past demographic or behavioural asymmetries between the sexes, for example in migration practices.

The essential raw material of genetic analysis, whether at the individual or population level, is variation. Genes and DNA sequences vary in many different ways and there are many methods to detect variation and a large battery of technical terms to describe it. The ultimate source of genetic variation is mutation, though at a local level variation may also be introduced into a population by migration. The pattern of variation present in any given generation of a population tends to be transmitted approximately to the next, with inter-generational variation in frequencies attributable possibly to natural selection (rarely demonstrable) or to chance (most likely to be important in smaller populations), as well as to migration. Sometimes inter-generational frequency fluctuation may end with the complete loss of a particular variant. Where a gene or DNA sequence exists in two or more fairly common variant forms, usually non-pathological, in a given population, the co-existence of those variants is described as a **polymorphism**.

Polymorphisms occur quite frequently in both genic and non-genic DNA, and can be highly informative about population inter-relationships. For example, where two populations both derive from the same ancestor population but are now separated such as by movement of one of them, one might expect, ceteris paribus, that diversification of an initially shared set of polymorphic gene frequencies will be progressive over time, except as constrained by natural selection. Often, especially in many of the non-genic DNA sequences, there are good grounds to doubt a major role for natural selection. In such cases most of all, population differences in the occurrence patterns and frequencies of polymorphic genes and DNA may be read as the traces left behind by past population origins and movements.

**References**


Robert Attenborough


Language families and the history of human migration

PETER BELLWOOD

Abstract

This paper looks at large scale language spread and replacement from the viewpoints of both recorded history and the comparative linguistic histories of major language families. It is suggested that the extents of language spread required to explain the distributions of the world’s most extensive language families, of foragers and food producers alike, would have required significant spread of native speaker communities. Language shift can explain small scale linguistic spread situations but not large language families, and elite dominance was not a major spreader of language families in pre-state times. The significance of reliance on fully domesticated and transportable plants and animals is also emphasised as a major factor behind the spreads of the large agriculturalist language families.

Prologue: Andy Pawley and the excitement of comparative historical linguistics

In early 1967 I crossed the world, courtesy of M.V. Achille Lauro, to join the Anthropology Department in Auckland University as a 23-year-old lecturer in archaeology. I knew a lot about Celts, Romans, Saxons and Vikings, but not much about Polynesians. Archaeology where I came from was normally linked closely to history or anthropology, but rarely to historical linguistics; indeed, my formal linguistic knowledge was probably limited to sources such as F.T. Wainwright’s excellent Archaeology, place names and history (Wainwright 1962). Of comparative linguistics as practised in the Austronesian world I knew nothing.

Within a very brief time after arrival in Auckland I was trying to understand how Polynesian archaeology, my passion at that time, related to Polynesian language history. I remember very well the little departmental tea room in an old wooden house on Symonds Street in Auckland, and the time when Andy raised the issue of why the majority of the peoples of New Guinea had retained their non-Austronesian languages, whereas the

---

1 An earlier version of this paper was presented at the NSF-funded Workshop on Alternative Approaches to Language Classification, organised by Philip Baldi, held in the Department of Linguistics at Stanford University from 17–19 July, 2007. I would like to thank Phil for the invitation to a very stimulating meeting.
original inhabitants of the islands of Southeast Asia to the west clearly had not. Why indeed? That issue still piques my curiosity today, although now the answer is clearer (Bellwood 2005:142–145; Pawley 2005:97–99).

Discussion with Andy, as well as with many others in Auckland, in particular Roger Green and Bruce Biggs, was central in developing my linguistic interests in those early days. I have kept in touch with Andy ever since, and in recent years both of us have researched on the same ANU campus in Canberra, albeit in different sections. Andy’s 2002 tour de force on Austronesian dispersal, published in a symposium organised and edited by Colin Renfrew and me, remains one of the clearest broad statements on this topic for palaeohistorians from all disciplinary backgrounds (Pawley 2002).

Introduction

The most eminent judgement to emerge from our global survey is that migrations of peoples, the first force in history to spread languages, dominates to this day. (Ostler 2005:534)

How can we explain the existence of any major language family, given that the ultimate origins and dispersals of none are documented historically, apart from the expansions of relatively recent subgroups and individual languages? My own feeling is that we cannot do this from the linguistics alone, but must take into account information from other disciplines, especially archaeology and human biology, including genetics. We also need a broad comparative perspective, one which takes account of how languages have actually spread in history, as recorded from the ancient empires onwards.

Most linguists with whom I have discussed the issue seem to believe that in order for any genuine (i.e. descent-based) language family to exist, the widespread observable patterns of shared retention and innovation necessitate a radiation of genealogically-related ancestral languages, commencing from an origin region and ultimately reaching the geographical boundaries of the family. Concepts of homeland and expansion are obviously crucial, and I have yet to meet any convincing claims that convergence of unrelated languages into a genealogical whole on a language family scale was ever a major factor, if indeed a factor at all in language family spread. Pidgins and creoles of course exist, but even in the event that a creolised Proto Austronesian language at 3000 BC ever spread and evolved into something on the scale of Austronesian in AD 1500, the descendants would not retain pidgin or creole status, but would exist as genealogically-related languages. Language spread would still be a non-negotiable requirement.

Genealogical relationship between different languages revolves around co-membership in language families such as Indo-European and Austronesian. The languages within such families share sets of common retentions from initial and reconstructable protolanguages. They form subgroups, defined by the sharing of innovations, that descend from reconstructable interstage languages at different temporal and geographical levels. Reconstruction is not always easy, since cognates can easily be replaced, and internal subgroup expansions within families can also replace other cousinly languages and subgroups, as must for instance have happened with all major linguistic spreads such as those of the Sinitic, Slavic and Romance languages in historical times. Any attempt to extract history from language families must always reckon with such problems. So-called macrofamilies do not feature in this paper, partly owing to the controversies surrounding their existence and membership, and partly because they are not essential elements in the debate about the origins of language families per se.
Language family existence requires foundation dispersal and progressive separation of speakers. Two important questions arise with respect to dispersal that are relevant for understanding broad-scale human prehistory from a linguistic perspective. Firstly, under what chronological circumstances did the initial spreads of specific language families occur? The answer for virtually all, perhaps not surprisingly, is in full and absolute prehistory. The initiations and early histories of none of these spreads are recorded in documentary history; all predate any historically recorded language expansions, rendering even Akkadian and Ancient Egyptian as little more than intriguing expansions on the margins of the vast Afroasiatic language family. Indo-European speech was established in Ireland and northern India long before the recording of any relevant mythology that can throw light on where Proto Indo-European was spoken and by what kind of society. Whole language families, as opposed to recent subgroups such as Romance, and single languages such as English, did not spread with historical states or conquerors.

Secondly, how were the foundation spreads of language families actually carried? The two extremes would be by migration of native speaker communities on the one hand, and by language shift through nodes in a chain with no speaker movement at all on the other. But extremes such as these rarely fit historical observations. Reality might generally have been something in-between, yet allowing prospects of allocating different relative significances to the polarities of speaker-movement versus language shift. This paper is concerned very much with that prospect.

Such questions of language family origin force us, on a region by region basis (there is no world wide formula), to consider the roles of hunters, farmers and conquerors, the significance of major economic and demographic changes in prehistory, and issues of chronology. Recovering the latter requires us to consider how to correlate reconstructed material lifestyles from comparative linguistics with the remains of such lifestyles in the radiocarbon-dated archaeological record. This paper thus limits itself to linguistics and archaeology, and focuses on the roles of systematic food production and demographic growth in the spreads of language families (Renfrew 1996). I do not attempt here to review the varying opinions about human migration presented by human geneticists, but I acknowledge that language spread can occur hand-in-hand with population admixture (Bellwood 2005).

How language families might have spread: lessons from the colonial past

Over twenty years ago, historian Alfred W. Crosby (1986) pointed out how intensive European colonization in temperate locations with few indigenous inhabitants — Argentina, Uruguay, USA, Canada, New Zealand, Australia — led to the establishment of what he termed ‘Neo-Europes’. Over 50 million European emigrants between 1820 and 1930 founded predominantly European populations speaking European languages (Spanish or English in most of the above cases). European-introduced epidemic diseases with their ‘devastating scythe-like effects’, their domesticated crops and animals, tough weeds, and high levels of human population fecundity in new-found fertile lands led to a tragedy for the indigenous populations of absolutely unparallelled proportions. Denevan (1992) has estimated native population declines of between 74 and 99 percent (the latter in Hispaniola) across the moving frontier of European conquest in the Americas. New Zealand in 1870 supported 250,000 Europeans and nine million sheep, but only 50,000 Maoris, who continued to dwindle in numbers from their pre-contact high of perhaps 100,000 or more until the population rebound of the twentieth century (Crosby 1986:265).
By the mid-nineteenth century, few Maori and Australian Aboriginal women were able to conceive owing to the devastating impact of introduced diseases, while European women in Australia in 1845 were giving birth to an average of 6.8 fairly healthy children (Vamplew 1987:55). Indigenous languages survived wherever indigenous population networks remained numerous enough to maintain them, but the overall impact of Spanish and British colonization in these regions was and still is absolutely obvious, especially in the language situation today.

Crosby went on to compare the Neo-Europes with locations that clearly were not Neo-Europes, mostly Old World tropical to equatorial locations with their own corpuses of diseases to which Europeans had no resistance, and all with dense populations living at high levels of social and political integration. The Middle East with North Africa, many regions of tropical Africa, and southern and eastern tropical Asia offer many examples (I omit the tropical Americas here because the linguistic outcome was so heavily structured by the massive loss of native populations due to introduced diseases, and the lack of any major reverse impact on the Europeans). European nations were unable to colonise very successfully in the Old World tropics, and once their empires had decayed, their languages faded from memory. Try speaking Dutch in Indonesia, French in Vietnam, or even English in many parts of India today. As Crosby (1986:63) rather amusingly commented on the attempts of the crusaders to conquer the Holy Land: ‘the conquerors, taken collectively, were like a lump of sugar presiding in a hot cup of tea’. In densely settled and culturally complex tropical regions, the boot was on the other foot. Here, the diseases hit the Europeans, not vice versa. European temperate crops would not produce well-filled ears of grain, pests abounded, and in most cases the European languages existed as little more than superficial lingua francas. It is true that some nations with immense linguistic diversity have decided to adopt a European language as a national language since World War II, but even in cases such as Papua New Guinea, with English and Motu as parliamentary languages, we must ask if the total population will one day speak only English, having abandoned all its native vernaculars? I doubt it, just as I doubt that the total population of India, where only an educated elite that comprises 2–3 percent of the total population is able to use English fluently (Britto 2003), will one day speak English and no other language.

The conclusion that can be drawn from the above is that single languages, when introduced into new territories, will normally only take hold on a permanent basis, as whole-population vernaculars, if they are imported in the mouths of substantial numbers of their native speakers. Native populations might undergo language shift, but only if the number of incoming speakers of the target language is sufficient to impose the required network of bilingualism for the shift to occur, or if they bring in potent diseases, as in Australasia and the New World, giving them the requisite demographic edge eventually to impose their language on the native population.

How language families might have spread: lessons from language history

Nicholas Ostler (2005) examines the history (or, in his words, the diachronic sociolinguistics) of the written languages of the world, from Egyptian and Akkadian of 2500 BC onwards to recent times. His conclusions, after 577 pages of detailed case studies, are that the languages of imperial administrations, and trade lingua francas, did not survive for very long after the systems that nourished them went into decline, unless there were implantations of very large numbers of native speakers of the imported languages concerned, or only a low density of aboriginal populations. As he observes (2005:275):
The cases where serious language change failed to follow on from conquests expose the hollowness of much military glory — the conquests in western Europe by Franks, Vandals and Visigoths, even the conquests in Britain by Romans and Normans.

Ostler discusses the rather feeble spreads, from a long term perspective, of a number of conquest-spread languages, including Ancient Egyptian outside Egypt, Aramaic (Persian Empire), Greek in Asia, Latin beyond the Romance-speaking areas (in which there was substantial colonisation by Latin speakers, since land was granted to Latin-speaking veterans), Germanic languages across post-Roman continental Europe, Arabic, Mongol (Genghis Khan spread lots of genes, but rather little linguistic heritage!), Norman French, and many more. Most of these examples fit the above quotation like a glove. Military conquest occurred, as history makes clear in no uncertain terms, but language spread only occurred, in general, in rough proportion to the degree of permanent population movement into the conquered areas. Ostler does not consider language families, and so overlaps little with the observations that I made in my own book First Farmers (Bellwood 2005). But the points about language spread that he makes for historical single languages are fundamental to what follows.

Even with languages adopted as national languages in modern times, first language speakers may be few. It is estimated that only 17 to 30 million out of a total population of 250 million Indonesians have Bahasa Indonesia as their first language (Ostler 2005:403), even though more than sixty years have passed since independence. Conquerors can also make it difficult for natives to learn their language — the Spanish in Latin America used native languages as lingua francas before 1770, and 52 percent of modern Paraguayans are still native Guarani speakers, with only 2 percent being bilingual in Guarani and Spanish (Ostler 2005:376). Language loyalty clearly matters to healthy and non-oppressed native populations.

Indeed, large numbers of indigenous language speakers, whether conquered or induced to learn imposed national languages, can give immunity against total language swamping from outside. Schooling (1990:123–124) has demonstrated this very clearly for New Caledonia, where native Austronesian languages are successfully resisting the domination of French:

Relationships are the glue that binds people together. If all the members of a community are individually bound to each other by a specific relationship, they form a solid block which will tend to stick together and conform to the same norms. This is DENSITY. If people are bound together by many different relationships, the link between them is like a many-stranded rope — very difficult to break. This is MULTIPLEXITY. People bound into a closely compacted social network have plenty of opportunity to be exposed to the standards and norms of that network. As long as there is a motive for remaining part of a network there is every reason to conform to its norms — to speak the language and do those things that are normal for that group. Even should there be a conscious desire to break free from such a network, it may be quite difficult to do so. Sanctions that can be applied to maintain conformity are often considerable.

In history, the Norman elite did not successfully impose French upon the medieval English, the English in turn did not successfully impose English on the Indians or the Malaysians. Language always serves as a badge of identity — people do not give up native vernaculars lightly. Even in cases of obligatory intermarriage between members of different language families, as in the celebrated case of the Vaupes region of north-western Amazonia, people still do not mix languages or undergo continuous language shift.
Instead, they maintain a situation of multilingualism, using their languages as relatively stable bases of ethnolinguistic identity (Aikhenvald 2002:23).

The history of Greek following the conquests of Alexander is one of the most striking examples of non-spread over the long term of a conqueror language. Greek colonies in central Asia incorporated core Greek speaking populations, but always in a sea of others speaking Indo-Aryan and Semitic languages — Greek was definitely the lump of sugar, not the tea. In Bactria, Greek declined in usage rapidly after the Greek garrison towns were overrun by Parthians, Sakas and Kushans, and by the early centuries AD only survived as a script used for the writing of Prakrit documents (Liu 2004). In Anatolia it survived through the Byzantine Empire, there to be replaced ultimately by Turkish after the fall of Byzantium in 1453. Why did Greek die so quickly? In central Asia, the reason is that there were never enough Greek speakers to push their language to the ‘tipping point’, compared to the much larger indigenous populations who surrounded them. Alexander and Greek philosophy were not enough, just as the Hindu pantheon was not enough to turn all the medieval Javanese into speakers of Sanskrit (or another Indic language), and Islam was not enough to convert the post-Majapahit Javanese into speakers of Arabic.

Some provisos can be discussed. Major language replacement might be assisted if the language is tied to a religion, and thereby released from the need for a migration of actual population. Yet Arabic only spread into regions where related Afroasiatic, and especially Semitic, languages were already spoken (Ostler 2005:97). Indeed, Arabic speaking populations already existed in parts of Jordan and Syria before the invasions (Nabataean Petra, Palmyra). It was also spread by conquering Arabic-speaking armies that set up garrison towns, and not entirely by language shift. The vast majority of the world’s Moslems today do not speak Arabic at all, except in the seventh century form used when reciting the Koran, preferring to maintain Iranian, Kurdish, Turkish, Punjabi, Swahili, Malay, Javanese and many other indigenous languages as their vernaculars. Neither did Christianity spread Greek, nor Judaism Hebrew. Indic languages derived from Sanskrit did not spread at all into Southeast Asia or China as vernaculars, despite the obvious successes of Hinduism and Buddhism amongst elite communities from the early centuries AD onwards across vast areas of eastern Asia.

Undoubtedly, language spread will also be lubricated if the language is written, standardised by a state authority, and not imposed under oppressive conditions by hated conquerors. Mass communication and school literacy help greatly as well, but none of these were features of Neolithic societies, for which nation states with single metropolitan languages are totally undemonstrated. The Neolithic/Formative world, within which the vast majority of the world’s great language families underwent their initial spreads, was a place of hunters, farmers, tribes, chiefs, occasional kings and towns, but not of imperial networks of administration. In such situations, Ostler’s observations about the pre-Roman Celtic migration into Galatia (central Anatolia) would have applied very forcefully:

(Galatian) was the language of a lineage. When its speakers moved, its domain would move with them, and if the community grew, so would the number of its speakers. If the community lost its identity, or its distinguishing customs, the language would disappear. (Ostler 2005:295)

The moral to be drawn from the above is that the long-distance spreads of the foundation languages that differentiated into the major existing language families of the world must have occurred with a continuous flow of native speakers, and not purely by processes of language shift. I certainly do not wish to deny the importance of language
shift in history — perhaps it occurred when Papuan speakers began to adopt Austronesian languages in the Bismarck Archipelago around 3000 years ago (Pawley 2002:267), just as when the descendants of Julius Caesar’s Gallic opponents eventually learnt Latin (or Proto French!) (Bauer 1996). But I wish to put language shift in perspective in terms of geographical scale. Although it occurs all around us today as the languages of nation states aggrandise and spread, most on-the-ground cases of language shift are small in terms of extent. For instance, Comrie (2002) discusses the shift in medieval Hungary from Slavic languages into Hungarian, and the shift in Azerbaijan from Caucasian languages into Turkic Azeri, but in both these cases the actual extents of the phenomena described are tiny in comparison to the overall extents of the Indo-European, Uralic and Altaic language families. Sims-Williams (1998:517) has suggested “… that language shift has been the rule rather than the exception’, but this statement ignores scale. It is one thing to be able to point to numerous small-scale situations of language shift in the recent and modern world, but another to make them significant for any discussion of continental-scale language family origins and spreads.

The discussion so far brings me to an intermediate conclusion. The foundation layers of the world’s major language families spread predominantly with populations of native speakers, and not by language shift, even though the latter will obviously have occurred in some small-scale situations. Convergence is not a viable hypothesis to explain the existence of any language family, allowing, of course, that dialects of a single language can reconverge to a degree through borrowing. But neither borrowing nor convergence can be regarded as processes of language dispersal.

**Dating the spreads of language families**

In the remainder of this paper I want to discuss how the spreads of language families that are so apparent to comparative linguists relate to actual chains of events in human prehistory. Several matters will concern us: geographical homeland and environment, cultural and especially economic background, and chronology. The focus, to reflect my own recent research interests, will be strongly on the role of food production in encouraging human dispersal within the past 10,000 years (Renfrew 1996, 2002; Diamond and Bellwood 2003; Bellwood 2001a, 2001b, 2005, 2009; Bellwood and Oxenham 2008). The discussion will be mainly structured around language families as the main taxonomic entities of interest, and I will focus on some of the world’s very large examples, particularly on those that have sufficient agricultural and pastoral content in their basal protolanguages to be considered as candidates for spread by early food producing populations.

To repeat, all of the major food producing language families (i.e. those for which a food producing vocabulary can be reconstructed from an early stage) had attained their AD 1500 distributional limits long before written history and any conquest empires. Secondly, all of the very widespread agriculturalist families would appear to have spread from homelands either within or very close to the archaeologically-defined homelands of food production (Middle East, west and Sahelian Africa, central China, New Guinea highlands, Mesoamerica, Andes and upper Amazonia, eastern woodlands), at least in terms of the majority homeland opinions of linguists (as discussed in Bellwood 2005:Ch.10, Figure 1). There may be argument, for instance, over whether early Indo-European languages spread from Anatolia or the Pontic steppes, but both of these regions are far closer to the Middle Eastern region of agricultural origin than are the Irish Republic, Norway or Bangladesh.
Major language families that have sufficiently reconstructed tree-like genealogies to allow homeland suggestions with some confidence include Indo-European, Bantu (in this case, within the broader Niger-Congo family), Dravidian, Tai, Austronesian, Uto-Aztecan (Hill 2007, in press), Arawakan and Tupian. Other families such as Afroasiatic, Austroasiatic and Sino-Tibetan do not have such clear genealogies, and the relatively rake-like and unrooted phylogenies often reconstructed for them allow less certainty (e.g. van Driem 2002 for Tibeto-Burman; Blench 2006 for Afroasiatic).

Figure 1: Suggested homelands for the major Old World language families, derived from interpretations of the linguistic literature. The Levant with Anatolia, West Africa, central East Asia (‘China’) and the New Guinea highlands are all recognised homeland regions for food production. See Bellwood (2005, 2008) for details.

These two factors — relative age of spread before the appearance of historical records, and substantial extent of spread — when combined with the observation that major language dispersal also required population movement — allow us to suggest that the most significant point of change in world prehistory within the agricultural latitudes could well have been the ‘transition’ from non-food production into food production (as suggested by Renfrew 1991, 1996). This was undeniably a major transition in the archaeological record, no matter how assiduously one tries to alter the semantic implications of that hoary old chestnut ‘the Neolithic Revolution’ (Gamble 2007). It could well turn out to have been the most significant fault line for language spread and replacement in prehistory.
But what was the actual structure of this fault line in the real-life histories of food producing populations? When did the agriculturalist expansions most likely occur, in a structural sense, along regional forager to farmer trajectories? Why were some far larger in extent than others? The answers lie partly in the definition of the term ‘food production’. In order to see the complexity of this, I have looked in a new publication (Bellwood 2009, and see references therein) at the sequence of early agricultural development in the area of the world best known in this regard — the Middle East, and especially the Levant and south-eastern Anatolia. The following list of stages is entirely archaeological and derives nothing from linguistics.

1. Early cultivation with purposeful planting during the Pre-Pottery Neolithic A (PPNA, 10,000 to 8500 BC) appears to have focused on morphologically wild cereals and legumes, harvested largely while still unripe in order to avoid the problems associated with shattering (many ethnographic populations harvested wild grasses at the unripe stage in Australia and India). Such practices are unable to select for a non-shattering phenotype in cereals and this type of pre-domestication cultivation probably did not induce very massive demographic growth in human populations, although the PPNA was a fairly widespread cultural phase in the Levant, including northern Iraq and south-eastern Turkey. This phase probably witnessed some increase in grain size, reduction of awns, and movement of cereals to locations in which they did not occur naturally, but complete agricultural dependence on domesticated crops was not yet achieved.

2. During the following Pre-Pottery Neolithic B, morphologically domesticated non-shattering cereals became gradually dominant after 8500 BC, with increasing harvesting of ripe grain, probably using flint sickles. At approximately the same time, animals were domesticated — sheep, goats, pigs and cattle — probably to maintain protein supplies in situations of increasing wild animal extirpation. The result of these developments was a considerable increase in demographic growth rates, as witnessed after 7500 BC by townships up to twelve hectares in size such as Abu Hureyra in Syria, Catalhöyük in Turkey and Ain Ghazal in Jordan. Such large townships must, of course, have been dependent on food production, otherwise they could not have existed. PPNB-related sites eventually spread over a larger area than PPNA — into Cyprus (by sea), central Turkey and through Iran to as far east as Pakistan by about 7500 BC.

3. After about 7000 BC, the later part of the PPNB and the following Pottery Neolithic were connected in the Levant and Anatolia with a widespread abandonment of many long-lived large settlements, to be replaced by smaller sites and increasing pastoralism. Perhaps this reflected a decline in agricultural productivity due to a combination of adverse climate and/or an over-exploitation of marginal environments. Whatever the answer, the advent of mixed farming exacted a rapid toll from the environment and doubtless set the stage for the next, and most significant, episode of expansion. An efflorescence of Neolithic expansion out of western Asia occurred; initially into western Anatolia and Greece by about 7000 BC and onwards into the Balkans (Ozdogan 2008), at around the same time into Pakistan, into

---

2 The heads of wild grasses typically shatter when ripe and scatter their grain. Domestication was based on the discovery of strains that did not shatter, suicidal for a wild grass but ideal for human cultivators who could bide their time and reap their crops all at once.
Turkmenia by 6000 BC, and by about 5500 BC into the Nile Valley. The expansion into Europe ultimately ended with the completion of the Neolithic colonisation of most of the temperate zone, apart from the northerly fringes and perhaps some regions of lingering maritime Mesolithic affluence, by about 4000 BC.

Thus, the agricultural dispersals out of the Middle East did not begin with a big bang precisely when the cultivation of plants is believed to have started. The Neolithic expansions into Egypt and Europe were the results of the development of a truly mixed farming (agropastoral) economy, not of early cereal cultivation alone. The passage of time between the first Natufian or PPNA experiments with cultivation and the beginning of the Greek Neolithic encompassed perhaps 4000 years, even 5000 years in the case of the oldest farming settlements in the Nile Valley.

Similar observations to those from the Levant are now apparent for China. Villages with plentiful evidence for the harvesting of rice and millets were well established in the alluvial plains of the Yellow and Yangzi rivers by 7000 BC, but dependence on truly domesticated non-shattering varieties of rice did not occur until after 5000 BC in the Yangzi region, perhaps slightly earlier for millets to the north (Fuller et al. 2007). This helps to explain why, given that fairly intensive pre-domestication cultivation of cereals was occurring so early in central China, the spread of rice cultivation into southern China and Southeast Asia only took place several millennia later. Dependence on fully-fledged food production, with consequent rapid demographic growth, did not develop overnight.

In this regard, the development of economies based on the reproduction of domesticated plants and animals, under human control and selected in terms of improved yield and transportability, becomes especially significant. Crosby’s (1986:89) concept of the ‘portmanteau biota’, of crops, animals and hitchhiking commensals carried by settlers and imposed in new landscapes, implies considerable levels of mobility and viability. Transportability would have been greater than with only wild species of plants and animals, which would have been harder to manage in captivity and more dependent for their continued reproduction on precise environmental variables. Kirch (1982) also discusses the propagation of ‘transported landscapes’ (after Edgar Anderson) in small and resource-poor islands in Polynesia, where perhaps the world’s clearest examples of this process can be found.

Beyond the portmanteau biota, other reasons for time-lags in the spreads of farming in early millennia are more purely geographical/environmental. In my First Farmers (Bellwood 2005), for instance, I drew attention to the marked delays in the spreads of food production between Mediterranean winter rainfall zones and monsoonal summer rainfall zones. For instance, food production required 4000 years to spread from west of the Indus into Peninsular India. In the opposite direction, food production did not spread into the Mediterranean climate regions of southern Africa, apart from Khoisan sheep herding, until modern times. Likewise, Uto-Aztecan maize farmers from the summer rainfall south-west were unable to carry or maintain their agricultural systems into winter rainfall California or the post-Fremont Great Basin. Advances of desert conditions also had a great deal to do with this. Instead, they were obliged to become hunter-gatherers, albeit with some degree of remnant agricultural technology (Hill 2002). Even with domesticated plants and animals, some environmental changes offered major barriers to early farmers.
Table 1: Likely origin regions and archaeological correlations for the initial spreads of major language families associated with food production

<table>
<thead>
<tr>
<th>Language family</th>
<th>Likely region of origin (from linguistic comparison)</th>
<th>Episodes and dates of correlative archaeological dispersal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indo-European</td>
<td>Eastern or central Anatolia</td>
<td>Western Anatolia to Thrace (and Black Sea?) 7000/6500 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>To Balkans 6300 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Northern Mediterranean 6200 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>To France (LBK) 5400-5000 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>To Denmark and UK 4000 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neolithic/Chalcolithic expansion via Iran into Xinjiang and Peninsular India (Tocharian and Indo-Iranian) between 6000 and 3500 BC</td>
</tr>
<tr>
<td>Dravidian</td>
<td>Lower Indus, Gujarat, or central Peninsular India</td>
<td>Contemporary with Indo-Iranian in NW South Asia and triggered by the same background conditions, thus after 3500 BC from Gujarat to Karnataka. Excluded from Ganges Basin by prior presences of IE and Munda (Austroasiatic) speakers.</td>
</tr>
<tr>
<td>Afroasiatic</td>
<td>Levant</td>
<td>PPNB expansion in Levant from 8000 BC, but to north Africa (caprines, Egyptian Neolithic) only after 7000 BC. (No cultural movement from Africa into the Levant is attested during the early Holocene)</td>
</tr>
<tr>
<td>Nilo-Saharan</td>
<td>Sahara/Sahel zone, during wetter early Holocene conditions</td>
<td>Spread of pottery and cattle management commencing c.8500 BC?</td>
</tr>
<tr>
<td>Bantu (subgroup of Niger-Congo)</td>
<td>Cameroon</td>
<td>Eastern Bantu spread with agriculture (and later with iron) between 1000 BC and AD 500</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>North-central China (Huang He basin)</td>
<td>6000 BC establishment of millet farming in Huang He basin, but spread of farming into Tibeto-Burman regions not attested until after 3000 BC</td>
</tr>
<tr>
<td>Language family</td>
<td>Likely region of origin (from linguistic comparison)</td>
<td>Episodes and dates of correlative archaeological dispersal</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Hmong-Mien</td>
<td>Middle Yangzi</td>
<td>Pre-domestication cultivation of rice 6000 BC, but HM linguistic spread is much younger owing to circumscription</td>
</tr>
<tr>
<td>‘Altaic’</td>
<td>Manchuria, Inner Mongolia, eastern steppes</td>
<td>Probably contemporary with Huang He developments — millet farming by 6000 BC, but dispersal mainly with pastoralism, especially for Turkic speakers</td>
</tr>
<tr>
<td>Tai</td>
<td>Guangxi/Guangdong/northern Vietnam</td>
<td>Origins with agricultural establishment in southern China, prior to second millennium BC, but most spread (especially into Thailand) much younger owing to homeland circumscription</td>
</tr>
<tr>
<td>Austronesian</td>
<td>Taiwan</td>
<td>Taiwan Neolithic standstill 3500 to 2200 BC, then dispersal (Malayo-Polynesian) in two rapid stages: (a) Batanes to Samoa second millennium BC (b) eastern Polynesia AD 800–1200</td>
</tr>
<tr>
<td>Austroasiatic</td>
<td>Assam-Yunnan? Ganges Basin? Southern China?</td>
<td>Spread of agricultural societies through mainland Southeast Asia and Ganges Basin commences c.3000 BC</td>
</tr>
<tr>
<td>Trans New Guinea</td>
<td>New Guinea Eastern Highlands</td>
<td>Spread of systematic agriculture with swamp drainage prior to third millennium BC</td>
</tr>
<tr>
<td>Uto-Aztecian</td>
<td>Valley of Mexico/central Mexico</td>
<td>Maize domestication by 4000 BC but spread to US Southwest during second millennium BC. Agriculture lost in Great Basin, and adopted by non-U.A speaking groups in Mogollon highlands</td>
</tr>
<tr>
<td>Otomanguean, Mayan, Mixe-Zoque</td>
<td>Mesoamerica (all circumscribed by the presences of each other, as well as by Uto-Aztecian to the north)</td>
<td>Most agricultural establishment is from 2000 BC, but some early agricultural activity from 4000 BC</td>
</tr>
<tr>
<td>Arawak</td>
<td>NW Amazonia?</td>
<td>Spread of agriculture into Amazonia unlikely to predate 2000 BC</td>
</tr>
<tr>
<td>Siouan, Iroquoian, Algonquian</td>
<td>Eastern woodlands — Ohio, Missouri and middle Mississippi basins?</td>
<td>Native seed crop agriculture by third millennium BC, but farming rejuvenated by dominance of maize after AD 500–1000</td>
</tr>
</tbody>
</table>
Conclusions

The reasoning behind this language-archaeology correlative exercise is based on the assumption that material culture and economic terms in reconstructed protovocabularies can be correlated in many instances with items in the material record recovered by archaeology. Crops, domesticated animals, and terms for activities associated with food producing technologies, were clearly definitive for Neolithic societies. Relating the linguistic records of agricultural language family spread and the archaeological records of farming spread is not circular reasoning, but our best way of using the data at hand for cross-disciplinary historical reconstruction.

If the proposals presented above are accepted, then certain historical conclusions follow. Many major language family dispersals represent demographic dispersal trajectories that can most parsimoniously be associated with movement of native speakers, at the same time not ruling out biological population admixture, as modelled for instance by Renfrew (2002) and Cavalli-Sforza (2002). In Table 1, I list major agriculturalist language families and their major subgroups, giving archaeological chronologies for the cultural expansions that can most likely be associated with them (Bellwood 2005). The principles of native speaker expansion are not restricted only to agriculturalist populations — they apply at all economic levels from Pleistocene and Holocene foragers to recent colonial settlers — and my focus on early agricultural populations here simply reflects my central research interest.

The suggestions made in this paper are based on a broad review of language family history and are intended to be points of discussion. Much of the supporting literature has been reviewed in Bellwood (2005), where dissenting views are also discussed, for instance over such specifics as the homelands of the Indo-European and Afroasiatic language families. The aim is not to make pontifical statements, but to encourage other scholars working in linguistics and archaeology to consider the points raised. History, in my view, deserves to be broad-scale and comparative, not ignoring the significance of local trajectories but also not uplifting them to the level where all grand syntheses are automatically rejected because somewhere there exists a minor contradiction. Having discussed these ideas with Andy over a period of 42 years I hope he will approve!

References


Friends, relatives, and enemies: the archaeology and history of interaction among Austronesian and non-Austronesian speakers in the western Solomons

PETER SHEPPARD, RICHARD WALTER AND KENNETH ROGA

In general, and particularly in the west, the Solomon Islands are an archipelago of inter-visible islands among which small fibre-glass and wooden dug-out canoes routinely move people and goods. Canoes from Maravari, a large village where we work on the south-east coast of Vella Lavella (Figure 1), travel on a daily basis along the coast of Vella and across to Ghizo, Kolombangara, Rannonga and the islands beyond. This interaction is reflected in language. Although the mother tongue of most people in Maravari is Bilua (a non-Austronesian language (Kauko Obata 2003), related, apparently at some time depth (Dunn et al. 2005), to the scattering of other non-Austronesian languages in the Solomons), many people are fluent in one or more of the Austronesian languages of their neighbours. This could be attributed in part to the high level of mobility and generally peaceful interaction which has characterised most of the last 100 years of Solomons history. Broadly speaking this is true and today villages may contain people originating from throughout the country. However, in the western Solomons high levels of regional interaction extend back well into prehistory. It seems probable that it has been a long time since geographical isolation has played a significant role in cultural and linguistic differentiation within the western Solomons. Interaction, both peaceful and violent, has been a driving force underlying the cultural and linguistic geography of the region. In this paper we will review what archaeology, oral history and history can tell us about relationships among the Austronesian (An) and non-Austronesian (NAn) speakers of the western Solomon Islands.

1 This research has been funded by the Royal Society of New Zealand Marsden fund. We would like to thank the people of Roviana and Vella Lavella (Maravari and Irigila villages) who have been so helpful to us over the last ten years. Leana hola and mata ouri. We also thank the many graduate students and other colleagues who worked with us in the western Solomons. Seline McNamee of the Department of Anthropology, University of Auckland produced Figure 1.
1 Linguistic and historical overview

Today 26 languages (Tryon and Hackman 1983) are extant in the western Solomons (New Georgia Group, plus Choiseul and Santa Isabel). Of these, 24 are Austronesian and, with the exception of Bughotu, belong to the Meso-Melanesian Linkage. Bilua (Vella Lavella) and Touo (Baniata area of eastern Rendova in the New Georgia Group) are NAn (Ross 2001, 1989:136). Formerly, other NAn languages were to be found in the region. Kazukuru for example, became extinct on western New Georgia in the early 20th century (Capell 1969; Waterhouse and Ray 1931). By the time it was recorded in Munda, on the coast of west New Georgia, it is probable that all speakers of Kazukuru had Roviana as their primary language as the interior Kazukuru area of western New Georgia had then been abandoned. Current research suggests (Dunn and Ross 2007) Kazukuru may have been an Austronesian language with considerable NAn admixture. It is also probable that a NAn language was spoken on the relatively large, uninhabited island of Tetapare which lies to the east of Rendova and appears to have had links to Baniata (Waterhouse 1927). The island of Ghizo was uninhabited at the end of the 19th century (Bennett 1985), but present-day Bilua speakers at Saeraghi village on the western tip of Ghizo claim to be descendents of people...
who were forced from the island by head-hunting in the 19th century. After the British established administration on the island and head-hunting ceased, these people returned to Ghizo at the start of the 20th century. These links hint at the former presence of NAn speakers on Ghizo and Tetepare but remain speculative.

It seems that the first languages spoken in the western Solomons were NAn, and a considerable number are found today on Bougainville, however only four (excluding languages with questionable NAn status in the Reef/Santa Cruz area) are found south of there. Although no archaeological sites of Late Pleistocene or Early Holocene age have been found in the western Solomons, the Kilu Cave on Buka (Wickler 2001) at the northern end of the Solomons dates back to circa 29,000 BP while the Vataluma Posovi Cave on Guadalcanal has deposits dating to the mid-Holocene (Roe 1993). Perhaps the best evidence of early settlement is the considerable amount of difference which exists among these NAn languages. Until recently it has been hard to determine relationships between them and the other NAn languages of the Solomons and Bougainville (Obata 2003; Ross 2001). Only recently Dunn et al. (2005) using phylogenetic methods, have proposed comprehensive relationships among these languages. Their work posits similarity among the NAn languages of the western and central Solomons and distinct separation from the NAn languages of Bougainville. This leads them to suggest a separation from a common ancestor of their Solomons and Bougainville language groups over 10,000 years ago (Dunn et al. 2005) when ancestral NAn speakers could migrate freely on Greater Bougainville (Figure 2), which linked Bougainville, Choiseul, Santa Isabel and the Florida Group during the low water period of the Late Pleistocene. We should note, of course, that none of the extant NAn languages are on islands which formed part of Greater Bougainville, and the inter-island distances between those islands and their neighbours, including Greater Bougainville would have been, during that time, much as they are today. Of particular interest is the positioning of the Solomons languages on a maximum parsimony tree (Dunn et al. 2005: Figure 4) between Bougainville and the Bismarck Archipelago. As Dunn et al. note this would contradict a simple model of geographic dispersal and branching from the Bismarcks and suggest a more complex history of divergence, perhaps reflecting, in our opinion, the relative geographic isolation of the Solomons languages from each other and the large contiguous NAn language distribution on Bougainville.

Austronesian languages are assumed to be associated with the expansion of the Lapita cultural tradition out of the Bismarck Archipelago some 3200 years ago. Although this is undoubtedly the case for Remote Oceania (Green 1991a), the situation in the main Solomons is potentially more complicated. It has long been argued that the distribution of the Meso-Melanesian Linkage represents a later intrusion of Austronesian speakers from the Bismarck Archipelago over the top of an earlier Lapita associated spread of Austronesian speakers. The descendents of these people speak Eastern Oceanic languages in the area east of the Tryon-Hackman line, which bisects the Solomons at the eastern tip of Santa Isabel and separates Southeast Solomonic and Meso-Melanesian (Lynch et al. 2002; Ross 1989:137). Whether we are dealing with multiple linguistic expansions into the western Solomons or not, it is clear that at some time, probably after circa 3200 BP, speakers of Austronesian languages began a long history of interaction with NAn speakers, resulting in the nearly complete geographical dominance of Austronesian in the region today.
2 Friends: early interaction

Although archaeologists do not ‘dig up’ language, it is reasonable to postulate that a major linguistic shift in a region will be marked at some level by a change in the archaeological record. It also seems clear that the spread of Austronesian languages is ultimately marked in part by the introduction of new food production systems and technology (Diamond and Bellwood 2003) although this need not discount the impact of earlier developments by NAn speakers (Allen 2003; Pawley 2005:98). In Remote Oceania, language, material culture, economy and biology made a generally coherent package which archaeologists have labelled Lapita. In Near Oceania this coherence need not exist given the evidence for much earlier settlement by NAn speakers. It would in fact be surprising if such coherence was common in the main Solomons. Archaeological evidence from the western Solomons, although limited, is now beginning to show a pattern of interaction, intrusion and integration as postulated by Roger Green (1991b) although the timing of this process seems to be later than proposed by what might be called ‘the classic wave of advance model’ which seems applicable to Remote Oceania and assumes an early Lapita movement out of the Bismarck archipelago and continuous settlement advancing throughout Near Oceania.

By 3100 BP makers of Lapita pottery were well established in the Reef/Santa Cruz islands of the Southeast Solomons (Green 2003) and by at least 3000 BP throughout the remainder of island Melanesia. Yet within the northern and western Solomon Islands including Bougainville and Buka there are no firmly dated ceramic bearing sites earlier than circa 2700 BP and, with the exception of two shards excavated by Specht (1969:194) on Buka, all sites containing Lapita dentate stamped pottery are inter-tidal lag deposits. Despite considerable research over the last 30 years in the western and northern Solomons
(Sheppard and Walter 2006), there is no record of early Lapita occupation in the region. On Buka inter-tidal reef flat deposits studied by Stephen Wickler (2001) have produced assemblages of Lapita ceramics. None of these are directly dated. Wickler (2001:241) suggests dates on coarse stylistic grounds — primarily on the percentage of dentate stamp decoration — in the range 3000 to 2800 BP (DJQ) and finishing circa 2300 to 2100 BP (DAF beach and inner reef). The DJQ site has over 56% (n=188) of decorated shards dentate stamped, but this falls sharply to 1.9% (n=77) at DAF (Wickler 2001:108), with dentate stamping replaced by increasing proportions of unbounded incision (Wickler 2001:112), punctuation and appliqué relief. The Late Lapita settlement pattern of inter-tidal stilt villages found in Buka extends into the western Solomons where over twenty such sites have been found (Sheppard and Walter 2006).

In Roviana Lagoon the work of our research team has located a considerable number of inter-tidal ceramic sites, and it is apparent that such sites are common throughout the western Solomons (Reeve 1989), wherever lagoon or sheltered fringing reef environments exist. In Roviana, Matthew Felgate (2001, 2003) has obtained the first absolute dates on these deposits by dating carbon in or on shards. Only a very small number of dentate stamped ceramics have been found at two Roviana sites (Honiavasa and Nusa Roviana). Most ceramics are decorated using incision or appliqué techniques. The dentate shards initially suggested a Late Lapita age, and this was strengthened by radiocarbon dates on a sherd from Panaivili (AA33504-charcoal) of 2130±90 BP and on an exterior smoke derived carbon deposit on a sherd from the Hoghoi site (NZA-1253) of 2619±45 BP (Felgate 2001:48). Both of these sites as described by Felgate (2003) contain rectilinear incised, punctate and pinched decoration as well as a variety of lip treatments including crenulated or horizontally deformed lips. Most obvious is the shift from complex pot forms in the apparently earlier Honiavasa site to simpler forms and often thinner wares in the post-Lapita sites (Felgate 2001:53). Although Felgate has argued that the absence of earlier Lapita sites dominated by dentate stamped ceramics may be a sampling or site visibility effect, on-going research over the last five years has rapidly increased the number of inter-tidal sites (Ghizo Island, Vella Lavella and Marovo Lagoon (Sheppard and Walter 2006); Kolombangara (Summerhayes and Scales 2005); and north New Georgia coast (Felgate pers. comm. 2003)) but failed to find any with more than a few pieces of dentate stamped ceramic or any terrestrial early ceramic sites.

The argument for a late Lapita associated spread into the western Solomons of a new settlement pattern and associated ceramic manufacture is strengthened by palynological evidence showing the first significant garden clearance circa 2600 BP. As part of the Roviana research, Sarah Grimes (2003) and John Dodson collected four pollen cores from swamps in the region. These cores show a pronounced and sustained rise in charcoal influx at circa 2600 BP, and that, in general, intense burning only developed in the later part of the 3rd millennium BP (Grimes 2003: Figures 4.5, 4.11, 4.17, 4.23). This correlates closely with the date on the Hoghoi site in Roviana, and is very similar to the pollen record from Guadalcanal where Haberle (1996) does not see significant evidence of slash and burn horticulture until circa 2200 BP.

It is conceivable that the appearance of this Late Lapita ceramic tradition, its associated settlement pattern, and the first appearance of sustained food production by slash and burn horticulture simply marks the diffusion of these new cultural forms amongst NA

\footnote{DJQ and DAF are the site codes for the sites referred to.}
speakers. Yet we know that at some point Austronesian speakers or at least Austronesian language began to spread throughout the region and it seems to be a reasonable hypothesis to suggest the concurrent spread of these new cultural forms and language circa 2700 BP. As we have argued elsewhere (Sheppard and Walter 2006), we propose that at that time there was a spread of Meso-Melanesian east from the Bismarcks out to the limits of archaeological and historic pottery distributions on Santa Isabel. To the east of the Tryon-Hackman line, at the eastern tip of Santa Isabel, there has never been any historic tradition of ceramic manufacture and, with the exception of a handful of plainware shards from rockshelters on Santa Ana (Black and Green 1975), there have been no archaeological discoveries of ceramics, other than those left by early Spanish colonisation attempts. This is despite considerable survey and excavation in the central and eastern Solomons, such as that of the Southeast Solomons Culture History Project (Green and Cresswell 1976; Yen 1982) and David Roe (1993) on Guadalcanal. Only in the remote south-east Solomons Reef/Santa Cruz Group, 400 km east of Ulawa, and points south of there do we find ceramics in numerous sites associated with the very early Lapita expansion into Remote Oceania which we have argued leap-frogged over the main Solomons. Subsequently at a point when ceramics were falling out of the cultural repertoire (circa 2000 BP), the descendents of these early colonists from somewhere in Remote Oceania in the greater south-east Solomons region extending down through Utupua and Vanikoro into Vanuatu, or at least their Austronesian language, would appear to have moved. This movement was to the east into the main Solomons, forming the modern distribution of Southeast Solomonic and the eastern most boundary, at the Tryon/Hackman Line, of Eastern Oceanic (Sheppard and Walter 2006).

Our thinking about the relationships among language, the archaeological record and biology is sharpened by consideration of our findings on Vella Lavella. At the northern end of the island, in the Irigila region, where fringing reefs and lagoon settings are common we have found considerable evidence of inter-tidal sites. Although we have yet to find any dentate ceramics, the ceramic assemblages are in all respects identical to those recovered from Roviana and elsewhere in the west and form part of this Late Lapita derived inter-tidal settlement tradition. The most parsimonious explanation for our finding on Vella is that NAn speakers have in this case adopted cultural traits from their neighbours. The alternative hypothesis, that NAn speakers replaced ceramic manufacturing Austronesian speakers at some date within the last two thousand years runs counter to the evidence that Bilua is a much older language which diverged from its closest NAn neighbours long before the inter-tidal tradition appeared. Along with ceramic production, and perhaps more fundamentally, it seems likely that at this time the people of Vella along with their NAn neighbours throughout the western Solomons adopted the new domesticates introduced from the west. Slight evidence may be found in linguistics for this development. For example, the Bilua term for ‘pig’ (m)bolo, is cognate with its Austronesian and NAn neighbours (Tryon and Hackman 1983:171).3

---

3 The word for ‘chicken’, kokorako, in Bilua is shared with all its Austronesian and NAn neighbours down to and including Lavukaleve on the Russell Islands, after which Savosavo (Savo Island) takes a form cognate with its Austronesian neighbours to the east. The word kokorako would appear to be pidgin and might suggest a late introduction of chickens into the main Solomon Islands.
3 Relatives: the genesis of Roviana

Today Roviana forms the largest group of Austronesian language speakers in the western Solomons, covering most of the south-west coast of New Georgia. In the 19th century Roviana Lagoon was the base of a powerful polity whose chiefs organised head-hunting raids throughout the western Solomons and east as far as Guadalcanal (Edge-Partington 1907). Although Roviana appears to be a quintessential Solomon Island Austronesian society, with core cultural features (e.g. mana, tapu) common throughout much of the Austronesian world, its own origin story states that the genesis of Roviana is founded upon an alliance of Austronesian and non-Austronesian speakers in the interior of north New Georgia. From here they moved together down to the coast and replaced the previous inhabitants (Koloi) (Aswani 2000). This interaction between adjacent populations from different language families was central to the formation of the Austronesian polity which came to dominate its neighbours and alter the linguistic history of New Georgia and the wider region.

The association of NAn speakers and land in western New Georgia is certain. The area in the interior behind the coastal Roviana town of Munda is called Kazukuru which is the name of the language with a NAn dominated vocabulary (Dunn and Ross 2007) recorded in Munda in the early 20th century (Capell 1969; Hall 1964; Waterhouse and Ray 1931). Whether Kazukuru is originally a NAn language with significant Austronesian admixture or vice versa, Roviana tradition describes the speakers as a linguistically and biologically distinct population (e.g. Hall 1964). Oral tradition records that Roviana chiefly genealogy begins with a woman called Roviana from Kazukuru whose descendent, the chief Ididubangara, formed an alliance between Kazukuru people and Austronesian speakers at a place called Bao in the interior of north New Georgia (Aswani 2000; Parker 1994). Together the combined tribes moved, circa 1600 AD, down to the island of Nusa Roviana which dominates Roviana lagoon. Relating this oral tradition to the archaeological or anthropological record is difficult but surprisingly not impossible. The fact that Roviana genealogy begins with a woman may be significant. Modern Roviana is not organised on a matrilineal basis but has a bilateral form of social organisation where descent groups (butubutu) trace rights on both sides. In the past, anthropologists and British colonial administrators (Allan 1957) trained to find ‘African-like’ lineal kinship systems (Scheffler 1965:35), have argued for an underlying matrilineal structure in the western Solomons. Modern anthropology, however, confirms the cognatic organisation which is common in the Solomons (Hviding 1996; Keesing 1970; Scheffler 2001) and throughout the Austronesian world. Where matrilineal organisation is clearly found in the western Solomons is in Vella Lavella where the landholding group (toutou) is traced through the maternal line and a man’s mother’s brother takes the lead in providing bride price and other social transactions. Angela Terrill (2003:3) has also noted that the NAn speaking people (Lavukaleve) of the Russell Group in the central Solomons have a strong matrilineal system of social organisation, seemingly very similar to that on Vella Lavella. Given the strong matrilineal focus among NAn speakers in Bougainville (Oliver 1993a, 1993b) this would seem to be characteristic, not only of ancestral NAn populations in Bougainville, as suggested by Oliver, but of NAn populations throughout the Solomons. Therefore, the woman Roviana may have been the apical female ancestor of a Kazukuru

---

4 Matrilineal descent is not, however, unknown among Austronesian speakers in the Solomons, with the nearest Austronesian speakers to Vella Lavella on the island of Rannonga practising such a system. The history of that system is unknown and requires further study.
matrilineage, and more generally, the early confusion over kinship and land tenure in the western Solomons (Allen 1957) may relate to confusion over current and relic NAn social organisation and its relationship or integration with Austronesian bilateral systems such as that in modern Roviana.

Our archaeological survey in the interior of New Georgia, behind Munda in the area called Bao, did in fact discover a large settlement complex located on a high ridge looking out to the north New Georgia coast. Despite considerable survey in the interior behind Roviana Lagoon, this complex at Bao remains unique. The large basalt platform and its associated features which sit at the eastern end of the ridge has given us our oldest date (1200 AD) on this early form of shrine platform which we find scattered in isolated positions along the interior ridges of Roviana. For Roviana people this shrine at Bao is the origin point of Roviana and this is in keeping with our radiocarbon chronology (Sheppard et al. 2004). This early form of shrine may have NAn origins (Hall 1964), but its widespread distribution in Roviana makes it impossible to clearly associate it with any language distribution and by 1200 AD it may have been widespread in New Georgia. We have not found anything exactly like it yet in Vella Lavella.

Although archaeology cannot provide us with any details of the on-going interaction between Austronesian and NAn speakers in Roviana, it does provide some important information on the outcome of this development which had both local and regional linguistic impacts. Oral tradition states that the chief Ididubangara came down from Bao to the island of Nusa Roviana. The island was historically the heavily populated fortified centre of Roviana from whence head-hunting raids were launched and where Roviana sheltered from attack by their neighbours, including during famous battles with warriors from Vella Lavella (Talasasa n.d.). Today the island contains abundant evidence of fortification (> 700 features), in the form of a series of high (3m+ in places) stone walls, platforms, terraces and shrines (hope) (Sheppard et al. 2000; Walter and Sheppard 2001). Dating of the fortification walls indicates construction at circa 1600 AD. A shrine complex containing numerous basalt columns and slabs derived from the mainland, which is traditionally associated with Ididubangara, dates to the same time period and is the earliest shrine on the island (Sheppard et al. 2004). Although the apparent support for the Roviana oral history is interesting, and supports the association with Kazukuru, of more importance for the history of the region is the fact that it is at this time we see the appearance in the Roviana sequence of the ancestral skull shrines and associated shell valuable complex which is characteristic of the head-hunting ritual system associated with historic Roviana.

4 Enemies: head-hunting and regional interaction

Head-hunting is one aspect of an ideological system which underwrites relationships between power, land and ritual in Roviana and throughout much of the western Solomons. In brief, heads acquired during raids are displayed in canoe houses where they materialise the *mana* (Dureau 2000; Keesing 1982) of the chiefs who organised the raids and own the canoes, thereby underwriting the authority of local leaders. *Mana* is also obtained from powerful ancestors whose skulls are placed in shrines where they are attended by priests under the direction of chiefs and elders. Shell valuables serve both as a ‘currency’ to reward warriors and fund head-hunting expeditions, as well as being symbols of power and ritual efficacy used as offerings on shrines and in other social transactions (Aswani and Sheppard 2003; Sheppard et al. 2004; Walter et al. 2004). Archaeologically this system is most evident in the shrines. These commonly have some form of elaborate stone
construction, a motu or offering fireplace, ancestral skulls and generally numerous shell valuables placed on or under the skulls (Figure 3). In Roviana this system appears circa 1600 AD when what was apparently an older, first fruits type of offering platform — lacking the skulls, shell valuables and offering fireplace — was transformed into the form known historically.

In Roviana these shrines are called hope. Similar forms along with the shell valuables are found throughout the western Solomons where they are also called sope (Choiseul, Santa Isabel,) or tabuna (Simbo, Rannonga) (Nagaoka 1999). At present we are still investigating the origin of this system, but it appears that the formation of the Roviana chiefdom played a central role in what was doubtless a regional process. Once head-hunting began, it spread rapidly, enveloping the region in a pattern of raid and counter-raid in which it was impossible to opt out. Along with the raiding the rest of the cultural complex, upon which it is founded, appears to have spread throughout the region.

Of particular interest to us in this paper is the spread of this Austronesian cultural complex founded on the concept of mana to Vella Lavella. Ancestral skull shrines are found throughout the island (Foanaota 1974) and they contain all the features found in Roviana, although their layout and contents may be influenced by the matrilineal social organisation onto which the cultural complex has been grafted. Small skull shrines are called sope while larger ones, which may contain a number of skull deposits relating to different matrilineages, are called matebangara. These are clearly Austronesian terms and would appear to be derived from Choiseul. On Choiseul the distinctive term mate bangara is used to refer to captives, usually young children, who are taken to ‘replace’ a particular person, often a recently-deceased leader. Through adoption the child might assume a leadership role in their adoptive descent group (Scheffler 1965:102). The terms used on Vella Lavella may reflect the incorporation of a number of different lineages into a shrine where one lineage had primary rights and other ‘foreign’ lineages were present under the auspices of the primary landholder. Shell valuables found on shrines include various kinds

---

**Figure 3:** Roviana skull shrine (hope) (Kindu Island, Vonavona Lagoon)
of bakisa (Roviana bakiha) which are associated with chiefly power. As in Roviana and Choiseul, the most valuable form of bakisa is manufactured from fossil giant clam shell (Tridacna sp.) found in upraised coral reef (makatea). Virtually all of the varieties of shell valuables found on shrines in Vella Lavella are replicated among their Austronesian neighbours. They were often used in combination to create nets which wrapped the skulls of important people (Figure 4), a practice which was widespread in the region (Barraud 1972). The date of the introduction of this cultural complex to Vella Lavella is still uncertain but the results of preliminary radiocarbon dating on shrines in the Maravari and Irigila regions indicates it was considerably later than in Roviana and possibly not much before the late 18th century AD.

Figure 4: Skull wrapped in shell valuables (Boroma site, south-east Vella Lavella)

Head-hunting clearly drove the introduction of Austronesian material culture and associated terms throughout the western Solomons, creating a degree of cultural homogeneity within the region which crossed linguistic boundaries. Unfortunately, the raiding also had significant impacts on the distribution of populations, especially in the 19th century when guns and the desire to obtain turtleshell for increasing trade with Europeans seem to have intensified the effects of head-hunting (McKinnon 1975). The small islands between Choiseul and Santa Isabel were rich turtle-hunting grounds and the adjacent islands provided sources of heads and captives for raiders from the New Georgia Group (Scheffler 1965:25). White (1991) reports that on Santa Isabel this period became known as the ‘time of flight’ when entire villages were exterminated. There were often not enough people left living to bury the dead. In the late 19th century, observers sailing along the coast noted that most of the eastern half of Santa Isabel seemed to be abandoned. The result of this violence can be seen today in the linguistic geography with only one language (Kia) spoken in the western half of Santa Isabel and seven in the eastern half of the island, where people retreated towards Bughotu and the protection of powerful leaders. Similarly, eastern
Choiseul supports only a number of dialects of Central-East Choiseul while three languages are pushed to the western end of the island (Tryon and Hackman 1983:23).

5 Relatives: integration of Austronesian speakers into Vella Lavella

It is unclear how much of an impact head-hunting had on Vella Lavella. In Vella Lavella people retreated to inland hamlets and defensive locations on the high ridges to shelter from raids by their neighbours; then however, we have not recorded any stories comparable to those reported by Geoff White (1991) for Santa Isabel. Roviana oral history recounts many stories of early raids and named wars with Vella Lavella, but it seems that in the 19th century raiding shifted towards Choiseul and Santa Isabel, and Vella may have been spared the full effects of the intensified 19th century warfare. Vella people did, however, raid their Austronesian neighbours from canoe houses located along the coast, and in addition to heads they returned with many captives. Those who were not ritually killed were eventually incorporated into society through adoption or marriage. As Debra McDougall (2000) has discussed for nearby Rannonga, societies throughout the western Solomons in the late 19th and early 20th century were very concerned about maintaining ‘tribes’ and descent groups in the face of the debilitating effects of head-hunting and European diseases. In Vella many foreigners were incorporated into local societies, both captives and people, usually women, who accidentally arrived on Vella or were brought in as marriage partners. This integration of Austronesians into Vella seems to have been a very important process throughout the island, at least from the mid-19th century.

In the village of Maravari today there are 11 named ‘tribes’ or matrilineages (toutou). Of these, five are described as foreign tribes with origins outside Vella Lavella. They are generally named for the location from which they came (Wagena, Wagena Island Choiseul; Ijo, Ghizo Island; Makavore, Choiseul; Sabe, Choiseul; Kolombangara, Kolombangara Island). Most of them seem to be descended from single ancestral women but the Sabe toutou is said to have come as a group from Choiseul. All of this movement appears to have occurred prior to the coming of the mission in 1902. One of the effects of this direct interaction with Austronesian speakers is likely to have been the introduction of words into Bilua. The study of this borrowing has been facilitated by the existence of an unpublished dictionary compiled by Methodist Missionaries at Vonunu on south-east Vella in about the early 1960s (Methodist Mission c.1960) and available in the University of Auckland Library Archives. Co-author of this paper Kenneth Roga, former Cultural Affairs Officer in the Western Province and current member of Provincial Parliament for north Rannonga, has examined the 2988 words in the Bilua dictionary. Kenneth is a native speaker of Rannonga, fluent in Roviana and well experienced in other Austronesian languages in the western Solomons. He has identified 10.3% of the words in the dictionary as comparable to forms he recognises from nearby Austronesian languages. Some of these may, of course, be Vella words adopted into Rannonga, but the vast majority seem to be Austronesian. Given that the early mission language was Roviana and the Roviana bible and hymnal were used on Vella, some of the language of the dictionary may be influenced by the mission context. However, much of it would appear to reflect earlier interaction. Many of the words relate to warfare, canoe technology, shell valuables and the general head-hunting context (Table 1). Other clusters of associated words are found related to the sea, fishing and house construction.
<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>aba</td>
<td>axe handle</td>
</tr>
<tr>
<td>bakisa</td>
<td>shell money</td>
</tr>
<tr>
<td>baga</td>
<td>cowrie shell</td>
</tr>
<tr>
<td>bakovara</td>
<td>barbed fighting spear</td>
</tr>
<tr>
<td>barava</td>
<td>a shell article</td>
</tr>
<tr>
<td>bareke</td>
<td>a shell or cane ring</td>
</tr>
<tr>
<td>bavi</td>
<td>crescent shaped shell ornament</td>
</tr>
<tr>
<td>beku</td>
<td>doll, image or idol</td>
</tr>
<tr>
<td>bokala</td>
<td>bow for arrow</td>
</tr>
<tr>
<td>bokolo</td>
<td>a clam shell armlet</td>
</tr>
<tr>
<td>bozo</td>
<td>body of a man killed</td>
</tr>
<tr>
<td>bugiri</td>
<td>a special kind of arrow</td>
</tr>
<tr>
<td>bulau</td>
<td>a small armlet</td>
</tr>
<tr>
<td>butu</td>
<td>bead or shell necklet</td>
</tr>
<tr>
<td>dala</td>
<td>a head ornament</td>
</tr>
<tr>
<td>era</td>
<td>burial place</td>
</tr>
<tr>
<td>ibiabala</td>
<td>a piece of money of lesser value than <em>bakisa</em></td>
</tr>
<tr>
<td>jelepande</td>
<td>place of worship</td>
</tr>
<tr>
<td>kabao</td>
<td>to return from head-hunting without heads</td>
</tr>
<tr>
<td>kana</td>
<td>enemy</td>
</tr>
<tr>
<td>keraova</td>
<td>shell ornament on canoe</td>
</tr>
<tr>
<td>kimo</td>
<td>top plank of canoe touched by paddles</td>
</tr>
<tr>
<td>lajono</td>
<td>a piece of wood used to keep canoe off ground</td>
</tr>
<tr>
<td>lalusa</td>
<td>final polishing of new canoe</td>
</tr>
<tr>
<td>lave</td>
<td>fibre used in canoe making</td>
</tr>
<tr>
<td>leviri</td>
<td>keel</td>
</tr>
<tr>
<td>liqomo</td>
<td>shell charm attached to shield of war leader</td>
</tr>
<tr>
<td>liqomo vovozo</td>
<td>the patron of war</td>
</tr>
<tr>
<td>lobusu</td>
<td>to draw sword from sheath</td>
</tr>
<tr>
<td>maja</td>
<td>a fighting axe</td>
</tr>
<tr>
<td>njujuluju</td>
<td>figure-head of war canoe</td>
</tr>
<tr>
<td>naqoto</td>
<td>top plank in war canoe</td>
</tr>
<tr>
<td>paele</td>
<td>canoe house</td>
</tr>
<tr>
<td>paka</td>
<td>gun</td>
</tr>
<tr>
<td>pauzuva</td>
<td>to adopt</td>
</tr>
<tr>
<td>pejao</td>
<td>feast on completion of a new canoe or house, formerly requiring human sacrifice</td>
</tr>
<tr>
<td>qata</td>
<td>a wooden spear or pole</td>
</tr>
<tr>
<td>qozato</td>
<td>an instrument for grinding the inside of money or armlet</td>
</tr>
<tr>
<td>sope pande</td>
<td>house for skulls</td>
</tr>
<tr>
<td>sopere</td>
<td>spear</td>
</tr>
<tr>
<td>tabu</td>
<td>sacred</td>
</tr>
<tr>
<td>taqutaquru</td>
<td>the regular beat of paddle on a gunwale of canoe</td>
</tr>
<tr>
<td>vozi</td>
<td>paddle</td>
</tr>
<tr>
<td>zerubule</td>
<td>cowrie shell used as charm on war canoe</td>
</tr>
</tbody>
</table>
6 Conclusion: archaeology and language — Austronesian or NAn
does it matter?

In the western Solomon Islands interaction between Austronesian and NAn speakers has
taken a variety of forms which have had significant outcomes for linguistic history and
geography and for culture history more generally. The earliest interaction resulted in the
transferral of new food producing technologies and food items, and probably new items of
material culture such as ceramics. These introductions, particularly the new domesticates,
would have been very attractive in an area where the land-based food resources would
have been limited. The strength of this new Austronesian cultural influence may have been
considerable and resulted in replacement of Nan, or at least the rapid introduction of
Austronesian into the comparatively small islands of the western Solomons circa 2700 BP.
This would seem to be attested to in part by the geography of Meso-Melanesian which
forms a coherent unit sharply separated from languages to the east of the Tryon-Hackman
line. Clearly two, regionally distinctive, historical processes have operated in the area. The
fact that this distribution is congruent with the archaeological distribution of ceramics and
the Late Lapita inter-tidal settlement pattern also suggests a relationship between the two
phenomena. The inter-tidal settlement pattern may have also facilitated occupation of
marginal areas on small off-shore islands not conducive to a gathering/hunting lifestyle,
allowing shared occupation of the landscape or at least a variety of lifestyles which may
have been favourable to the maintenance of a number of linguistic traditions.

Following the early period, interaction resulted in the near extinction of NAn languages
throughout the region and the development of a great diversity of Austronesian languages
forming dialect chains in the major islands (Ross 1988). Something of the nature of this
process is indicated by the history of Roviana where we can initially see the formation of
alliances between NAn and Austronesian speakers followed by the complete assimilation
of the NAn speakers. In this case it appears that the expansion of Roviana, which was
based on the elaboration of a cultural system rooted in key Austronesian concepts, resulted
in the co-option or destruction of all of its near neighbours including Kazukuru with which
it was closely integrated. Similar processes playing out repeatedly over thousands of years
have seen the gradual replacement of NAn throughout the region. It seems unlikely,
however, that these historical processes have often been as dominating as that occurring in
Roviana over the last 500 years. Roviana grew to become the largest language group in the
region and its activities, or activities generated by the head-hunting complex, had major
impacts on its neighbours, reshaping significant parts of the language geography of the
western Solomons. After the coming of the mission the adoption of Roviana as a lingua
franca has also facilitated the spread of the language in a more peaceful fashion. The
development of great Austronesian linguistic diversity and dialect chains suggest that for
much of prehistory there were no equivalent disruptions to the gradual erosion of NAn by
Austronesian.

Why NAn languages were maintained in some areas like Vella Lavella remains a
mystery. It may simply be an historical accident, although this suggests that from an
evolutionary perspective the overall selective advantage of speaking Austronesian over
NAn was slight. It may also be that given a few hundred more years NAn would have been
replaced throughout the region. On Vella the head-hunting complex appears to have
resulted in the rapid introduction of Austronesian words, concepts and people. Today
chiefs in Marvari complain that the old language is being lost and that their children’s
vocabulary is full of English.
Should archaeologists be concerned about the prehistory of NAn and Austronesian? Some would argue that as we can never make any direct association between language and the archaeological record, archaeology and historical linguistics should travel completely separate paths, meeting occasionally to exchange notes. In Island Melanesia there has been considerable debate over the association between Lapita and Austronesian. This has resulted in a variety of models of both the origins of Lapita and the settlement of Near and Remote Oceania. Clearly in Near Oceania any model which demotes interaction between earlier and later linguistic and/or biological populations will have limited utility. It may be that interaction between NAn and Austronesian speakers was one of the key drivers in both the development of Lapita and its expansion, but how could we ever investigate this archaeologically? Perhaps we should maintain archaeological purity and stay away from linguistics. Yet if our goal is to write culture history, and we acknowledge that this may not be the primary goal of many archaeologists, to abandon historical linguistics is to throw away a powerful source of data. Linguistics can provide us with models of historical and geographical patterns of cultural variation which in some ways are analytically superior to the efforts of archaeologists. The result may shed light where archaeology is absent and generate useful hypotheses for testing by archaeological means. In the western Solomons linguistics provides a fairly strong argument for at least early Holocene settlement of the region (Dunn et al. 2005), something missing so far from the archaeological record. In the Late Holocene, linguistics indicates the presence of two important historical processes which have generated the pattern demarcated by the Tryon-Hackman line. Careful evaluation of the archaeological record for the Solomons shows that the linguistic pattern is matched in the archaeological record. The archaeological evidence, supported by palynology, provides a chronology for these processes and a new interpretation of how the linguistic pattern developed. Furthermore our hypothesis provides room, within the Solomon Island context, for both a ‘fast train’ model (Oppenheimer 2004), which introduces Austronesian and a coherent Lapita culture directly to Remote Oceania, while allowing for an ‘entangled’ (Terrell et al. 2001) development in Near Oceania (Sheppard and Walter 2006). We emphasise, however, the need to consider developments within particular regions rather than use simple coarse models which attempt to summarise very large regions and long time periods.

Clearly linguistics can provide very useful models of large scale historical geography and processes (Kirch and Green 2001; Pawley and Green 1973; Ross 1998) but does it matter to the archaeological record if the makers of that record were NAn or Austronesian speakers? In Roviana we cannot discriminate between the Austronesian and NAn archaeological record. One of our goals in working on Vella Lavella is to see what a NAn record might look like. As we have discussed above, the Bilua speakers of Vella rapidly incorporated the culture of their Austronesian neighbours, most spectacularly the head-hunting complex. Explaining the general development of this record on Vella will be little different than explaining it anywhere in the western Solomons. However, it seems likely that explanation of variation in this general phenomenon will benefit from our knowledge of the fact that the Vella people have participated in a different historical tradition, as signalled by language, than their neighbours. For example, our preliminary results suggest that the presence of distinct, lineage-related skull deposits within matebangara on Vella is related to the presence of matrilineal organisation on Vella, something which may have been common to NAn speakers within the Solomons. Explaining this variation in the record would not be possible without considering the prior population history which
established the conditions to which any introduction had to adapt. Language is a powerful signal of history, and archaeology in the Pacific owes much to the collaboration with linguists and especially the pioneering work and friendship of Andy Pawley (e.g. Pawley and Ross 1993). Such a major contribution cannot be easily dismissed.

References


Black, Steven and Roger Green, 1975, Radiocarbon dates from the British Solomon Islands to December 1973. Auckland: University of Auckland, Department of Anthropology.


Methodist Mission, c.1960, Bilua dictionary. Photocopy held in Special Collections, University of Auckland Library, Auckland, NZ.


Friends, relatives, and enemies


Talasasa, John, n.d., The fighting at Roviana. MS held in Special Collections, [George G. Carter] Papers, 1960s–1970s, University of Auckland Library. Auckland, NZ.


‘I was so much older then, I’m younger than that now’: why the dates keep changing for the spread of Austronesian languages

MATTHEW SPRIGGS

Linguists have never had a reliable means of dating the break-up of the various early subgroups of Island Southeast Asian and Pacific languages independent of the archaeological signatures associated with the Austronesian expansion (Figures 1 and 2). Glottochronology has come and (largely) gone, but occasionally linguists present a ‘rule of thumb’ and unquantified version of what amounts to the same thing. This is used to say that the differences between nominated successor protolanguages would have been like the difference between languages whose divergence times are known through historical records. The Romance languages are those most often being thought of. On the other hand, the cautionary tale of the very conservative Icelandic language sometimes gets a salutary mention as well.

Pawley (2004:251) has made a very important observation concerning the Austronesian languages in his suggestion that models of linguistic change applied to continental areas may be inappropriate. He notes that Austronesian is the only major language family predominantly spoken on islands, and many of these islands are isolated by hundreds or even thousands of kilometres. Why should one expect similar rates of linguistic change to those found in areas such as Europe when this is the case? Pawley and Ross (1995:65–66) made a similar point in relation to dialect chain models. Such models are perhaps more suited to island situations than the ‘wave models’ used in Indo-European linguistics.

Love them or hate them, it is clearly the archaeologists who have the key to dating Austronesian linguistic change. For areas in Remote Oceania (Andy Pawley’s term for a concept developed by Roger Green: Pawley 2004:269, fn.5) beyond the main Solomons chain it is a very powerful master key indeed. This is because the initial inhabitants of this

---

1 This paper was written in 2006 while shifting between Canberra, Vanuatu and the United Kingdom on research funded in part by the Australian Research Council. I would particularly like to thank my parents Joyce and Frank Spriggs for their toleration and hospitality while I was writing the bulk of the text in Petersfield, and Dr Una Strand Víðarsdóttir for providing the space and time for me to finish it at Barnard Castle. The latter also provided helpful grammatical correction. John Bowden and two anonymous referees provided further useful suggestions. Andy Pawley has been a consistent proponent of the need to integrate archaeological and linguistic data in the Pacific region, and I have valued his comradeship and support in the struggle to do so.
area were Austronesian speakers. One can follow the distinctive pottery-using culture of Lapita at the base of archaeological sequences in the western parts of Remote Oceania back to its ‘homeland’ in the Bismarck Archipelago, and beyond to its antecedent Island Southeast Asian neolithic cultures. There is a high level of cultural continuity displayed by this cultural horizon. Its distinctiveness from what came before in regional cultural sequences allows a rare matching of archaeological and linguistic reconstructions. This then gives us the ability to date the Austronesian spread and the consequent break-up of its various linguistic stages.

Archaeology is able to provide a far clearer picture for this part of the Asia-Pacific region than it can for the spread of Indo-European languages. There are simply too many comparable spreads of archaeological horizons across Europe at too many different times for us ever to be sure whether Indo-European spread at the beginning of the neolithic as Renfrew and others have argued (Renfrew 1987), or whether it is exclusively a Bronze Age phenomenon as is also cogently argued (Mallory 1989).

There are on-going problems of matching the archaeology and the linguistics in our region too. These are in part to do with an undeniable fact: it is more straightforward and much less costly to describe a language to a degree useful for comparative purposes than it is to define a 10,000 or even 3000 years or less archaeological sequence for any comparable area. Many islands that have quite decent dictionaries and grammars (and often the entire Bible translated into their constituent language or languages) have yet to receive the undoubted benefits of a visit by archaeologists. Unsuspecting speakers of a range of languages can be enticed to leave their shores to share their knowledge with linguists in the metropolitan centres; but the archaeology is not at all transportable. The archaeologists have to go to the mountain, it won’t come to them. It is usually fairly obvious if there are speakers of a particular language present in an area. It is much harder to find the key archaeological site or sites that will establish a detailed archaeological sequence for the same locale. Needles and haystacks often come to mind during such searches.

1 ‘It’s getting better all the time’

This situation also implies, of course, that things will inevitably improve with time. Just as a lot of new comparative data are generated every year by diligent linguists to prove grist to the historical linguistic mill, so too the dogged persistence of the archaeologists tends to pay off in the end. As well as better sample coverage as new regions are systematically researched, there are ongoing improvements in dating methods for archaeological materials. Radiocarbon dates quoted in the earlier literature tend to fall by the wayside as we understand more about the suitability of particular materials for dating and as the accompanying contextual information is refined. Many important sites have recently been re-dated using the latest refinements in dating technology.

When applied to Pleistocene sites beyond 10,000 year BP (‘before present’) the general trend has been for human occupation of regions to be found to be older than first thought, but for a range of reasons to be noted here this is not the case for key sites thought to be associated with Austronesian spread. In fact it is the opposite. Yesterday’s 4000 year-old Lapita site is today’s 3500 or 3300 year-old one, Yesterday’s earliest settlement date for the Marquesas of 100 BC is today’s 600–1000 AD settlement date, and the much-quoted 800 AD date for the earliest settlement of New Zealand is now looking more like 1250 AD. Why is this?
Figure 1: The Austronesian language family and major subgroups (adapted from Ross et al. 2003:xx).

Figure 2: Geographic limits of historically known Oceanic speakers and distribution of Lapita sites (adapted from Ross et al. 2003:xxi).
A major problem, long-recognised but not able to be acted upon due to limitations of expertise and funding, is the ‘old wood’ problem. Radiocarbon dating does not tell you when a fire was lit on a beach by its first visitors, it tells you when that part of the tree died that was being fed into the fire. Given a tree of any size, the interior of the trunk could have been growing some hundreds of years before the outermost parts or the small branches. There are some long-lived species in the region that present this problem to the archaeologist. For dating analyses the solution is to select only either quick-growing species or small twigs that represent recent growth. Such selection is not always possible, however, if there is no palaeobotanical expertise available to make the necessary species identification.

A related old wood problem is that of the burning of driftwood, some of which can come from long-lived North American species that have drifted out to sea and across the oceans. When people first reached particular Pacific Islands they would have had to climb over large piles of such driftwood to get to the shore. When recognised and dated, even modern driftwood can sometimes be over a thousand years old, as Kenneth Emory and Yosi Sinoto discovered at Ka Lae (South Point) on Hawaii Island nearly 50 years ago (Emory and Sinoto 1969:4). A detailed study of wood charcoal in Kaho’olawe Island (Hawaii) archaeological sites revealed that a significant percentage of all wood being burned there was exotic driftwood with the potential for extensive inbuilt age (Murakami 1987).

Just as sieve size has come down so that small fishbones and other remains can be recovered from the archaeological deposits, so have the thicknesses of excavation units been reduced. This gives less chance for materials of a considerable time range to become mixed together. Much smaller samples are now needed for radiocarbon dating, particularly with the development of Accelerator Mass Spectrometry (AMS) dating; this again means less chance of mixing materials of different ages. There is now a much greater understanding of the stratigraphy of sites and of the many natural and cultural factors that can cause site disturbance, mixing deposits of different ages. Significant in this regard has been a growing recognition that much of the visible stratigraphy of cave sites, particularly the beautiful ‘layer-cake’ appearance revealed in some rockshelter excavations, is entirely post-depositional (Spriggs 1999). It comes from chemical alteration of the sediments after they have been deposited and buried. The consequence is that the original culturally-altered stratigraphy, with its pits and postholes disturbing and churning lower deposits, has become homogenised, masking such early disturbance features. Many of our sites are significantly disturbed and mixed and there is nothing we can do about that except to try and evaluate and test for such factors after excavation. Careful excavation cannot compensate for this fact.

The radiocarbon laboratories too have improved over time. Some of the early ones were shoestring operations that doubtless meant well but produced erroneous ages, sometimes due to factors such as unstable modern standards to which they were attempting to calibrate the results, or insufficient pretreatment of samples to remove contaminants. Materials such as human and animal bone that were once thought to provide reliable ages are now known to be very tricky to date. Determinations run on these materials before the 1990s have had to be rejected, and samples re-dated (where possible) using modern pretreatment methods.

It was a combination of all these factors and more — the above listing has not attempted to be exhaustive (see Spriggs 1989 and 1999; Spriggs and Anderson 1993 for further
— that produced a run of radiocarbon dates now realised to be too early from a range of key sites in Island Southeast Asia and the Pacific during the 1950s through the 1970s. Simple chance may have been involved in this pattern as well, as in any probabilistic procedure; the roll of the radiocarbon dice. Being the first dates from particular regions or island groups, these tended to be seized upon by the keener linguists and their archaeological congeners. Now long rejected by archaeologists, the news about such dates often seems to take some time to reach linguistic colleagues. An ‘old archaeology’ problem has thus been created for linguists seeking to match the two data sets.

This has seriously discombobulated some linguists. An archaeologist might point out, say, that Fiji was not settled at 3500 BP as first appeared to be the case archaeologically but only about 3000 BP. What might appear at first sight to be a reasonable response from the linguist could be to suggest that archaeologists have not yet found the earliest site and one of that age will eventually come to light; needles and haystacks again. However, for Remote Oceania at least, there is a constant signature of earliest occupation in the presence in sites of large numbers of bones of extinct birds and other vertebrates, that time and time again have been shown to go extinct within about 200 years or so of initial human settlement of an island. A combination of human predation, habitat disturbance, predation and harassment by introduced rats, pigs and dogs, and avian diseases introduced by the domestic chicken seem to have done for the Pacific birds in what is the greatest vertebrate extinction event ever in terms of species, dinosaurs notwithstanding (Steadman 1999:376). If the bird fauna of a site are what is there today, or was there at least immediately before European contact, then you have not found a site of initial settlement.

A range of other indicators can be used to back up the initial settlement status of sites: further evidence of human environmental impact in the geomorphic and pollen records, the very large size of shell fish in early sites, the presence of large amounts of turtle bone, and so on (Weisler 1998:77–82). In combination they show us when humans first turned up, or at least when agricultural populations with domestic animals showed up. We do not necessarily get these signatures of human impact in Near Oceania at initial settlement way back in the Pleistocene, but we do see important environmental changes with agriculture in that region.

A certain kind of linguist might think he or she finds support here for another proposition, that the features of the languages say of Vanuatu suggest that there was an earlier Papuan-speaking population there which influenced linguistic developments. They weren’t agriculturalists so the archaeologists might not find the evidence of human impact that they have been expecting. Maybe not, but enough work has been done in Vanuatu — much of it targeted at the issue of pre-Austronesian settlement — to be reasonably confident that there was no-one in that archipelago (or in Fiji, or in New Caledonia for that matter) before about 3100 BP.

Evidence of Pleistocene settlement in Near Oceania usually comes from cave sites. They form good concentration points for the accumulation of settlement debris generated by mobile hunter-gatherers. In New Ireland six cave sites have been excavated by archaeologists and all of them contain pre-3000 BP material. On Buka in the northern Solomons two cave sites have been excavated. Both of them contain pre-3000 BP occupation; the same is true on Manus Island in the Admiralties (see Spriggs 1997:47, 73). In contrast, across the archipelago of Vanuatu somewhere approaching 30 cave sites have
been dug and none have even Lapita age occupation in them. Cultural sequences generally begin several hundred years later than the Lapita period (Bedford 2006; Bedford et al. 1998).

Areas with very high uplift rates, preserving ancient near-shoreline environments which have otherwise disappeared with rising sea-levels prior to 6000 BP, have been especially investigated in Vanuatu with the aim of finding pre-Lapita settlement — but with no result. Archaeologists really do seem to have established the boundary between pre-Austronesian and Austronesian human settlement, and it is where Pawley and Green (1973) predicted it would be — the Near Oceania/Remote Oceania boundary at the end of the main Solomons chain. If there are indeed Papuan elements in Vanuatu languages they must reflect contacts after first Lapita settlement; it was the Papuan speakers who would have been the later migrants (cf. Pawley 2006).

Linguistic opinion too has of course changed over the last 40 years, so perhaps archaeologists should not be too defensive over changing their minds during the same period. The standard view of Austronesian subgrouping only really firmed up in the late 1960s to early 1970s, through the work of scholars such as Pawley in the Oceanic region and Blust for the Indo-Malaysian parts of the story (see references in Pawley 2004:252–253). Synthesis involving archaeological dates for linguistic stages was only really possible in the early 1970s and can be seen in particular in two early papers by Pawley and Green (1973) and Shutler and Marck (1975).²

I will now examine the use of archaeology in dating various linguistic stages in Pawley and Green’s (1973) paper to illustrate the points made earlier about subsequent developments in archaeology that have sometimes contributed to the linguists’ ‘old archaeology’ problem.

2 Proto Austronesian to Proto Oceanic

In 1973 several factors led Pawley and Green to argue for a pre-Lapita spread of Austronesian languages into the Pacific as far at least as Vanuatu and New Caledonia. One was the siren song of glottochronology, in the form of Dyen’s (1965, 1971) classification of Austronesian languages. Dyen argued that Proto Austronesian must have broken up by 5000 BP (3000 BC), based on lexicostatistical calculations. It was also considered that Proto Oceanic must have broken up at least a few hundred years afterwards. Taking a conservative estimate derived from Dyen’s figures, Pawley and Green suggested a date in excess of 4000 BP for this linguistic stage (1973:27, 54). These authors were well aware of the criticisms made of glottochronology and lexicostatistics, but felt that there was archaeological evidence to back up the dates derived from Dyen’s work.

The archaeological back-up consisted of evidence from several dated sites that at the time seemed supportive of Dyen’s conclusions, and indeed seemed to push the dates back even further: over 6000 BP for Proto Austronesian and 5000 BP for Proto Oceanic. From

² There are earlier examples where linguists appealed to archaeological evidence provided by radiocarbon dates as an interpretive aid, but these occurred before any generally-agreed model of Austronesian subgrouping had been developed. An example is Grace (1964:366) who cites a radiocarbon date from New Caledonia of ‘847 BC’ in support of his contention — which turned out in fact to be correct — that: ‘I find it hard to believe that the Austronesian spread into eastern Melanesia was not completed by 1000 BC’ (ibid.). This date comes from his rule of thumb assertion that the development of the innovations forming Proto Polynesian would have taken about 1000 years, and his noting of the date of ‘122 BC’ from the Marquesas — taken from Suggs (1961) — as an anchor point for the break-up of Proto Polynesian.
north and west to south and east the adduced evidence was as follows. Current views on these dates are given immediately after each point.

1. Pawley and Green (1973:fn.20) refer to breaking news of archaeological discoveries in an unidentified conference paper by Solheim. This must refer to the research of Spoehr in Zamboanga and Sulu in the southern Philippines. Solheim (1975:24) interpreted Spoehr (1973:11, Figs. 117d and 180) to have shown that: ‘Red-slipped pottery with impressed circles inlaid with lime may be as early as 6650±180 C14 years B.P. on Sanga Sanga Island in the Sulu Archipelago […] Also from this site come shell adzes and gouges similar to those associated with the early Lapita pottery’. Close examination of Spoehr’s text and of the publication of subsequent re-excavation of the Sanga Sanga rockshelter (Ronquillo et al. 1993) convince me that this site is hopelessly disturbed. The visible stratigraphy is entirely post-depositional in origin, to do with chemical alterations of the sediments subsequent to their deposition. Currently acceptable dates for early pottery in this part of Island Southeast Asia all postdate 4000 BP (see below).

2. Pottery in East Timor was dated to 5500–4500 BP (Glover 1969, 1986). Again, site disturbance is now seen to be the reason for this apparent association. No significant archaeology took place in East Timor during the Indonesian occupation 1975–1999 to contradict these initial results. The latest research there suggests a date in the range 3800–3600 BP is much more realistic for the beginning of neolithic, Austronesian occupation (O’Connor and Veth 2005; other sources cited in Spriggs 2006).

3. There was thought to be early evidence for pigs and agriculture in New Guinea: ‘Numerous pig bones appear in two of the Highlands sites at about 3,000 BC [5000 BP], while extensive forest clearance took place at about that date in the Western Highlands’ (Pawley and Green 1973:6, citing Allen 1972 and Powell 1970; cf. Blust 1976). At that time it was accepted that several important Oceanic crops along with the pig, dog and chicken were introduced by Austronesian speakers. The date of their introduction, with evidence for the agricultural systems that would have gone along with them, would therefore be the date of the spread of Oceanic Austronesian languages. More recent discoveries and reinterpretation of the origin of the New Guinea domesticated crops now suggest that the region was an independent centre of agricultural development preceding by several millennia any Austronesian contacts (Denham et al. 2003, 2004). The early dates for pig have not been confirmed; again unrecognised site disturbance is the most likely explanation (Spriggs 1996a, 1996b). Even had these ages been confirmed, the eventual claims for pig at 10,000 BP were later recognised to relate to putative events long prior to Austronesian spread (cf. Blust 1976:26–27).

4. Pawley and Green (1973:7–8) state that: ‘[…] between 5,000 and 3,000 years ago quadrangular and planilateral adzes appear in New Guinea … Quadrangular adzes were almost certainly part of the Proto-Austronesian tool-kit … Planilateral adzes are also associated with early assemblages in Remote Oceania, and may have been introduced into New Guinea by Austronesian speakers.’ Whether there are really fully-polished stone adzes of these or any other shapes in New Guinea at such ages remains in doubt. All such finds were made early in the development of archaeology there and have not been backed up by more recent research (cf. Spriggs 1996a).
remain skeptical as to the claimed dated associations. Additionally, in relation to the presence of fully-ground stone adzes in Near Oceania, Green (2000) has since concluded that the planilaterial adzes were a Lapita culture innovation, rather than having been introduced from Island Southeast Asia.

5. ‘The New Caledonia sequence begins with an aceramic tradition which is the oldest yet discovered for any part of Remote Oceania […] This tradition was seemingly present between 3000 and 1000 BC and perhaps before then (much earlier dates have been obtained but their interpretation is uncertain)’ (Pawley and Green 1973:11). The evidence was that from the mysterious ‘tumuli’ or mounds, but was not fully accepted by the authors even at that time. These mounds are now generally agreed to be natural rather than cultural features (Green 1988).

6. It was also believed that initial occupation of southern Vanuatu had been aceramic, with the dates from 2500 BP onwards being seen as suggesting a significantly earlier date for occupation of that archipelago. Other artefacts recovered from this and later time periods in southern Vanuatu were seen as characteristically Austronesian in association (Pawley and Green 1973:12–13). In 1984 I was able to report the discovery of Lapita pottery in southern Vanuatu, and subsequent work has confirmed that the initial occupation of that region was part of the Lapita spread at about 3100 BP (Bedford 2006; Spriggs 1984).

In 1973 the lack of pottery in what were thought to be the earliest Austronesian sites suggested that there was an (at least in part) aceramic Austronesian spread in the western Pacific as far as New Caledonia around 5000 BP, associated with the spread and break-up of Oceanic Austronesian. It was noted that the earliest Lapita pottery dates were those from the Reefs-Santa Cruz Group in Remote Oceania at about 3250 BP (Pawley and Green 1973:19). It was assumed that dates from the putative Lapita homeland in the Bismarck Archipelago would turn out to be a few hundred years earlier, but such dates were unlikely to go back to the time of initial Oceanic spread suggested by lexicostatistical analysis and seemingly confirmed by early aceramic occupation of archipelagoes such as New Caledonia. Dates for pottery in Vanuatu or the south-east Solomon islands were seen as: ‘[…] far too late to account for the kind of linguistic diversity found in these regions. If diversification began by 1,000 BC in the relatively homogeneous Central Pacific region [Fiji, Tonga and Samoa], it must have begun considerably earlier in the island groups further west’ (Pawley and Green 1973:49).

Pawley and Green were, however, somewhat equivocal over their interpretation of an early pre-Lapita Austronesian spread. They realised that Proto Central Pacific did appear to have a Lapita association in Fiji, Tonga and Samoa, and that Proto Central Pacific had its most likely antecedents in North/Central Vanuatu languages:

The most likely explanations are (a) that some North Hebridean-Central Pacific speakers adopted a Lapita culture, (b) an immigrant Lapita community adopted a North Hebridean-Central Pacific language. If earlier dates for Lapita in the New Hebrides [Vanuatu] are recovered, dates of the order of 2,000–1,500 B.C., some of the difficulties in associating Lapita with Proto-North Hebridean-Central Pacific would, however, disappear. (Pawley and Green 1973:49–50).

Since 1973 there have been a few additional furphies, such as Egloff’s first date for early Lapita in Mussau in the Bismarck Archipelago of 3900 BP (Bafmatuk et al. 1980), and claims for 5000 BP pottery, pig and betel chewing near the north coast of the Sepik
region of New Guinea. This latter claim was believed to attest to an earlier Austronesian spread (Swadling 1996 and 1997; Swadling et al. 1991). These cases are not generally accepted (Lilley 2004:94; Spriggs 2001). Indeed, the early betel nut claims have since been withdrawn (Fairbairn and Swadling 2005).

2.1 Proto Central Pacific and Polynesian languages

Pawley and Green (1973) seemed to be on stronger ground in relating archaeology and linguistics in Fiji and Polynesia. Much more research had taken place in this region by the early 1970s, with a lot more radiocarbon-dated sites able to be considered. However, exactly similar problems as have become apparent further to the west can be seen to be operating in this region too: ‘old wood’, dodgy radiocarbon dating results, questionable associations between dated samples and cultural materials, and problems in the dating and interpretation of pollen cores. At least in Polynesia, however, these problems have had a much more public airing than in areas to the west. The issues, even if unresolved, are considerably clearer to practitioners. For eastern Polynesia, research programs specifically dedicated to sorting out the chronology of occupation at key sites have been instituted within the last decade and have overturned much of what we thought we knew about the colonisation of eastern Polynesia.

If we go back to 1973, again there were a series of radiocarbon dates from research carried out in the 1960s that set the tone for subsequent interpretations. These gave results often several hundred years older than now seems likely for initial settlement of a range of archipelagoes and islands. Notable among them were:

1. the date for the Natunuku Lapita site in Fiji of 3240±100 BP;
2. for eastern Polynesia, Suggs’ (1961) dates for the Marquesas Islands that suggested initial settlement some 200–100 BC; and
3. a date for Easter Island suggesting settlement there around 300–400 AD (Smith 1961).

The settlement of all the main archipelagoes of Polynesia was thought to have been completed by about 800 AD, the generally accepted date for settlement of New Zealand.

Pawley and Green showed a remarkably prescient suspicion of several of these early dates; the presence of pottery at some sites in the Marquesas suggested to them that the initial phase of Marquesan settlement had not yet been found. They considered that a date around 200 BC to 200 AD would one day be confirmed for the first eastern Polynesian landfall, with the Marquesas unlikely to be the first island group reached by settlers from western Polynesia (Pawley and Green 1973:22).

3 The emperor’s new raiment

Spriggs and Anderson (1993) were the first to question the by-then conventional model of east Polynesian settlement dates, as Anderson (1991) had already convincingly done for New Zealand. They suggested a model of later settlement of east Polynesia than usually considered. Their argument was that ever-earlier claims for settlement were being accepted uncritically and could not be sustained by the evidence. Prominent among these were the continuous settlement model of Irwin (1992) that suggested no pause between the
settlement of west and east Polynesia, and Kirch’s interpretation of his recent excavations and palaeoenvironmental research on Mangaia in the Cook Islands (Kirch et al. 1991). Irwin’s model was also criticised on linguistic grounds by Pawley (1996). The latter suggested that there must be a substantial pause after initial settlement of western Polynesia for the many innovations in Proto Polynesian to have developed. These innovations constituted the most strongly-marked Austronesian subgroup between Taiwan and the Tuamotus and implied a significant period of development before eastern Polynesia was settled.

Since publication of Spriggs and Anderson (1993) the trend has been for further research to suggest even later dates for settlement of particular archipelagoes. Indeed there is some reason to believe that this modelling has got a little out of hand. Some of the latest models for east Polynesia would seem seriously at odds with linguistic subgrouping arguments, unless much more rapid linguistic change is to be admitted than currently accepted. There have been some subtle but significant changes on the linguistic subgrouping side within Polynesian as well, particularly the clarification of the position of Mangarevan and the postulation (admittedly controversial: see later discussion) of a south-eastern Polynesian subgroup (Fisher 2000:2001). The previous isolate of Easter Island is now argued to be a part of this subgroup.

Spriggs and Anderson (1993) suggested that the central Polynesian islands (minimally Tahiti and the Marquesas) were settled in the interval 300–600 AD, Hawaii at about 600 AD, Easter Island around 600–900 AD, and New Zealand at about 1200 AD. Although it seemed possible that the Cook Islands were settled early, given their location, these two authors could find no convincing evidence at that time for any occupation prior to about 1000 AD. There was a forthright defence of early settlement models, but no new data, from Kirch and Ellison (1994), but Kirch now seems broadly to have adopted the Spriggs and Anderson model (compare Kirch 2000).

New data have come from various research projects, some specifically designed to test amongst the early and late settlement models for east Polynesia. Many of the data are conveniently summarised in Anderson (2002) who reports on the re-dating of some of the classic sites in French Polynesia, such as Ha’atuatau in the Marquesas — where Suggs had originally produced the 200–100 BC radiocarbon dates — and Vaito’otia-Fa’ahia on Huahine and Motu Paeao on Maupiti in the Tahitian group. Anderson’s conclusion on the basis of this new evidence is that:

French Polynesia may not have been colonized earlier than about AD 900–1000. That in turn implies a younger chronology for East Polynesia as a whole […] Instead of supporting the general conclusion of colonization beginning in East Polynesia at 300–600 AD, reached by Spriggs and Anderson (1993), much less at periods up to 500 BC envisaged by Irwin (1992), Kirch and Ellison (1994) and others, the case which must now be answered proposes initial settlement in Central East Polynesia no earlier than late in the 1st millennium AD. (Anderson 2002:251; see also Anderson and Sinoto (2002) for further data from Hane in the Marquesas.)

Further pertinent dates have come from Easter Island where Steadman et al. (1994) conducted new excavations at Anakena that suggested initial occupation at this key settlement site began about 800–900 AD. A later analysis of the radiocarbon data from Easter Island suggested to Martinsson-Wallin and Crockford (2002) that the island was settled sometime between 800–1200 AD. Weisler and colleagues have been conducting new excavations and reinterpreting old finds from other parts of south-eastern Polynesia,
particularly Henderson, Mangareva and Pitcairn Islands, to provide firm evidence for occupation in these groups by 900–1000 AD (Conte and Kirch 2005; Green and Weisler 2000, 2002; Weisler 1998). As these islands are the obvious stepping-stones between Central Polynesia and Easter Island, such dates imply a rapid colonisation of this whole area.

Most recently, Kennett et al. (2006) have produced an initial settlement date of AD1200 for isolated Rapa Island in the Australs Group. The Rapans speak a language derived from Mangareva after that group was ‘invaded’ by Southeast Marquesan speakers. Initial material evidence for such an intrusion occurs in the period around 1200 AD on Mangareva (Green and Weisler 2002:233) but could probably be pushed back by about 50 years or so. This date is significant in that it suggests that the recent estimate of 1200 AD for settlement of Easter Island (Hunt and Lipo 2006), derived from new excavations at Anakena and a re-analysis of the corpus of radiocarbon dates from the island, is most likely too young. If Mangareva is the proximate source for Easter Island settlement then on linguistic grounds it must have taken place prior to Marquesan influence on the Mangarevan language. This is because Easter Island shows no trace of such influence. Hunt and Lipo (2006:1605) hedge their bets a bit towards the end of their paper: ‘Additional radiocarbon dates will likely change the probability distribution and could reveal colonisation of Rapa Nui sometime slightly earlier than 1200 AD (e.g. circa 1050 to 1150 AD)’. One might add, or even earlier of course!

The further fringes of Eastern Polynesia, Hawaiian and New Zealand have also produced new radiocarbon determinations since the early 1990s. Hawaii had long been claimed to have been settled around 3–400 AD on the basis of two sites: Pu‘u Ali‘i (H1) on Hawaii Island and Bellows Dune (O18) on O‘ahu (Kirch 1985). Re-dating of both sites has failed to confirm the earlier determinations: Dye (1992) has re-dated Pu‘u Ali‘i to around 1400 AD and Tuggle and Spriggs (2001) dated further samples from Bellows Dune to suggest that occupation began around 700–1100 AD, most probably at about 1000 AD. Further analyses of the radiocarbon corpus and other evidence for Hawaii have suggested 700–900 AD as the likely date of first settlement (Masse and Tuggle 1998). The re-dating of the earliest phase site of Wairau Bar in New Zealand, often quoted as proving settlement at around 800 AD, has now produced dates for initial occupation at around 1250–1300 AD (Higham et al. 1999; cf. Hogg et al. 2003).

Anderson’s (2002) dating estimates for French Polynesia are unlikely to provide sufficient time for the linguistic changes that distinguish the various subgroups of East Polynesian to have taken place, especially if one assumes that the Tahiti-Marquesas area was where Proto East Polynesian was spoken. However, such problems are minor when one compares the shortest of the short chronologies, that of Pearthree and Di Piazza (2003).

These authors wish to pack the entire colonisation of East Polynesia into the period 1100–1400 AD. They believe this period represents the time necessary for the establishment of communities from a homeland in the southern Cooks out as far as New Zealand on one hand and Easter Island on the other. This is largely based on an analysis of

---

3 The Cook Islands consist of very small islands that may have been initially ignored for permanent settlement in favour of the much larger and more densely-packed islands of the Tahitian and (to a lesser extent) Marquesan chains if they were all discovered at the same time by scouts. There is certainly no mother-daughter relationship evident between Cook Islands’ and Tahitian languages as might be expected if the Cooks were the original east Polynesian homeland.
dates for ‘Archaic East Polynesian’ (AEP) material culture, which they believe represents the culture of the earliest inhabitants. There is no space to go into this issue here, but the lack of this culture in Hawaii, for instance, suggests that it is in fact a later, albeit widespread, development.\footnote{Pearthree and Di Piazza do note that the lack of all but one of their supposedly diagnostic Archaic East Polynesian (AEP) artefacts in Hawaii suggests ‘a culture history somewhat different from the rest of East Polynesia’ (2003:330). They offer no interpretation of this difference however. Only in the southern Cooks are they unsure that AEP represents the earliest occupation of the islands, but they do believe that AEP culture arose there, on the basis of geographical propinquity to western Polynesia and because of the presence of abundant pearl shell in the northern Cooks, used for some of the most typical AEP artefact forms (Pearthree and di Piazza 2003:334).} It should postdate the initial occupation of archipelagoes such as Hawaii and those in south-eastern Polynesia, including Easter Island where its signal is weak (Kirch 1986). In the Cook Islands and New Zealand AEP may well be the basal culture, supporting the evidence that these island groups were settled later rather than earlier in the settlement sequence.

Green and Weisler (2002), on the other hand, are perhaps a bit guilty of trying to force all the east Polynesian dates earlier on somewhat flimsy grounds, in an attempt to accommodate the linguistic evidence. An outlier date on bird bone from Henderson of 1295±50 BP (Weisler 1998:84) is claimed as making a 700 AD date of initial occupation of that island, and by extension the whole region from Mangareva to Easter Island, a ‘distinct possibility’. All other dates begin about 800–900 AD for this area. Green and Weisler also rely on the ‘semi-domesticated’ coconut date from Mo’orea — used to suggest occupation in the Tahitian group began around 600–800 AD — to get the central Polynesia dates back towards the mid-first millennium AD (Lepofsky et al. 1992 [cited in Green 1996:217]). However, as pointed out by Anderson, there is a logical flaw in this argument: the coconut could have drifted from western Polynesia where there were domesticated forms available from about 3000 BP (2002:247).

Green and Weisler subtly nudge even this date back further, as they do the well-attested dates for Hawaii and south-eastern Polynesia. The Society Islands’ date now ‘predates A.D. 600’, while they consider there is ‘strong support’ for colonisation of Hawaii and south-eastern Polynesia ‘in the interval A.D. 700–800’. It is further adjudged on the Tahitian evidence that ‘in the Southern Cook Islands, it should be even earlier’ (2002:236). Expansion into south-eastern Polynesia is again nudged possibly earlier, back to 600–800 AD in the conclusion of their paper (Green and Weisler 2002:237). The primacy of the Cook Islands in east Polynesian settlement at earlier than 600 AD relies on Kirch and Ellison’s (1994) interpretation of the palynological data from Mangaia, an interpretation which has been vigorously challenged by Anderson (1994; 1995).\footnote{Although I have pointed out some areas of disagreement over dates with Green and Weisler (2002), it should be pointed out that their general model, the first archaeological presentation to incorporate the linguistic insights of Fischer (2000, 2001) into East Polynesian subgrouping, is extremely significant for any consideration of the archaeology of the region. I am aware that some linguists are extremely dubious on strict linguistic grounds of the reality of Fischer’s Southeastern Polynesian subgroup (Marcr 2002, Ruter 2002), but as Marcr at least seems to allow, it still remains a — and arguably the most — plausible hypothesis to explain both the linguistic and archaeological evidence that we have for the area.}
Figure 3: Schematic diagram of the diversification of Austronesian languages (from Ross et al. 2003:7).
4 ‘I read the news today, oh boy’: a summary of the current estimates for the spread of Austronesian languages

Figure 3 presents a schema for the diversification of Austronesian languages, to be discussed in the following sections. The rationale for the dates I provide below has been given elsewhere in relation to Island Southeast Asia and the western Pacific (Spriggs 2003, revised in Spriggs 2006, and references cited therein). For eastern Polynesia it depends upon consideration of recent evidence cited above. I will tie the dates to the relevant linguistic subgroups or language stages, as discussed by Pawley (2004). Further detail on particular linguistic situations and their archaeological correlates is very well covered in Green (1999). He provides a subtle analysis of the processes of linguistic change in the region which I fully support, except (as will be obvious from the discussion above) for the dates given in relation to east Polynesian settlement.

4.1 Proto Austronesian (PAn)

The situation in Taiwan is that the neolithic begins at about 6000–5500 BP and develops from a mainland Chinese source probably in the Pearl River Delta of Guangdong Province (Liu 2006; Tsang 2006). There is extensive occupation throughout Taiwan prior to neolithic settlement to areas further to the south. This is witnessed over a considerable period by the development of the several first-order subgroups of Austronesian within the island.

4.2 Proto Malayo-Polynesian (PMP)

PMP is the ancestor of all Austronesian languages outside Taiwan, and so where it was spoken should also be the area for earliest neolithic settlement in that region. As the other nine primary subgroups of Austronesian were spoken on Taiwan, one would probably not be far wrong in suggesting that PMP developed when a colonising group moved out of that island. Tsang (2006) has noted that red-slipped pottery and other material culture similar to the earliest neolithic assemblages found in northern Luzon in the Philippines occur on the east coast of Taiwan from about 4000 BP and not before. This would therefore suggest a limiting date for further neolithic expansion.

The logical first step in such an expansion was to the Batanes Islands, between Taiwan and Luzon. The most recent evidence from there would suggest that its settlement date may need to be pushed back perhaps as far as 4500 BP on the basis of links with even earlier Taiwanese pottery styles (Peter Bellwood pers. comm. 2006; cf. Bellwood and Dizon 2005). For the time being, however, I will take a conservative line and note that there is no evidence for neolithic occupation beyond the Batanes until around 3800 BP. Over what area the dialects of PMP would have been spoken is unknown. Given the archaeological evidence for early sites and the boundaries of subsequent linguistic subgroups one might suggest it covered the Philippines, Sulawesi and perhaps adjacent parts of eastern Borneo (Pawley 2004:261). PMP broke up into Proto Central/Eastern Malayo Polynesian and the languages of the Western Malayo-Polynesian area.
4.3 Proto Central/Eastern Malayo-Polynesian (PCEMP)

This subgroup would have been formed when Austronesian speakers spread beyond Sulawesi, perhaps initially continuing to the south into the Lesser Sunda Chain, and then moving predominantly anti-clockwise along it and up into central Maluku. Neolithic occupation of East Timor is first attested in the period 3800–3600 BP, perhaps nearer to the latter date. PCEMP broke up into Proto Central and Proto Eastern Malayo-Polynesian.

4.4 Western Malayo-Polynesian (WMP)

Western Malayo-Polynesian is simply a shorthand for all the Malayo-Polynesian languages which did not descend from PCEMP. It is a residual category and not an innovation-defined subgroup. Some quite extensive subgroups within it are now being identified, however. One of the most recent to be put forward is Malayo-Sumbawan, incorporating languages of Malayic-speaking areas, Chamic, Balinese-Sasak-Sumbawa and Sundanese and Madurese. However, this subgroup is not strongly marked, suggesting a rapid spread of Austronesian speakers into the area after the break-up of PMP (Adelaar 2005). This would have been part of a rapid spread of languages through the Philippines, Borneo, Sumatra, Java and Bali, forming a rake-like pattern of relatively small subgroups. There has also clearly been a subsequent linguistic levelling, for instance in the Philippines, that obscures the original pattern of language spread. The earliest acceptable dates in the western Borneo, Java and Sumatra regions start around 3500 BP, but the situation in Borneo is still somewhat confused.

Bellwood (2005:6) has recently rejected his earlier suggestion (1997:237–238) of a pre-Austronesian neolithic spread encompassing western Borneo, Sumatra and parts of Java. He linked this to a movement of Austroasiatic speakers out of the Malay Peninsula. He now sees the cord-marked pottery of this region as deriving from the Fine Corded Ware of Taiwan of about 4500 BP, although the trail of this pottery does not extend back along the presumed migration route as far as northern Luzon. Given how little neolithic archaeology has been carried out in Sumatra and Java, I find this rejection to be somewhat premature for the region as a whole. Sumatra is directly adjacent to areas of the Malay Peninsula whose neolithic quite clearly does derive from movements of Austroasiatic speakers down the Peninsula. The first modern excavations of neolithic material in Sumatra have produced pottery linked by its excavators to mainland Southeast Asian traditions (Guillaud 2006; Simanjuntak and Forestier 2004). There is a clear change in much of western Indonesia from early cord-marked to later red-slipped pottery that is much more in the mainstream of Island Southeast Asian Austronesian neolithic cultures.

In postulating this view I take a somewhat different position on the spread of these languages to Pawley (2004:261, 268) who sees the migration stream as moving south-east into Maluku from either the southern Philippines or northern Sulawesi, and then spreading in a clockwise direction through Maluku and west along the Lesser Sunda chain. A spread initially to the Lesser Sundas and subsequently into central Maluku fits better with the currently available archaeological dates — admittedly few in this entire area — and provides a route for the spread of the Pacific clade of pigs along the Lesser Sundas from an ultimately mainland Southeast Asian origin, into Maluku and subsequently out into the Pacific (Larsen et al. 2007).

Anderson (2005) has recently taken up Bellwood’s original formulation, calling the two spreads ‘neolithic I’ (Austroasiatic-associated) and ‘neolithic II’ (Austronesian). His further suggestions, phrased disarmingly as questions, are original but in the end unconvincing: ‘[…] could neolithic I have dispersed sufficiently far eastward to have been connected to possible mid-Holocene pottery in northern New Guinea, and possibly in
An Austroasiatic, neolithic substratum, at least in Sumatra and Java, would also provide a route for the introduction of the Pacific clade of pigs, to which nearly all Pacific island pigs belong, from mainland Southeast Asia along a southern route. This route avoided Taiwan and the Philippines whose pigs have different genetic origins (Larson et al., 2007). A spread of pigs down through Sumatra and Java to be taken up by speakers of PCEMP in the Lesser Sunda Islands, the conduit for further spread of the Pacific clade out into the Pacific, must have taken place prior to about 3600–3500 BP. This date range is required in order to fit with the likely dates for PCEMP and the subsequent interstage language Proto Eastern Malayo-Polynesian (PEMP). It was the break-up of the latter, with the movement of Austronesian speakers along the north New Guinea coast to the Bismarck Archipelago that goes with the introduction of these pigs into the western Pacific.

Beyond Island Southeast Asia, WMP languages are also spoken in the Mariana Islands and Palau in western Micronesia. Not surprisingly the initial dates of occupation for these archipelagoes are comparable to other sites in the WMP area beyond the Philippines and Sulawesi. There is an extensive sequence of dates from the Marianas beginning about 3500 BP, and the earliest acceptable dates from Palau begin at about 3400 BP. They must represent two separate migrations from the WMP area. Dates for WMP languages on the mainland of Southeast Asia are later still and seem to reflect subsequent terminal neolithic or early Metal Age expansions of population into Vietnam and the southern parts of the Malay Peninsula, beginning about 2500 BP.

4.5 Proto Eastern Malayo-Polynesian (PEMP) and Central Malayo-Polynesian (CMP)

The impression from the available neolithic dates is that there is something of a pause between occupation of the Philippines, Sulawesi and the Lesser Sundas and that of areas of northern Maluku just to the east. We might estimate 3500–3300 BP for settlement of northern Maluku and the Cenderawasih Bay area of western New Guinea, to fit with the beginnings of Lapita further east in the Bismarck Archipelago at about 3300 BP (see below). Linguistically it is thought that there was primacy of settlement in Cenderawasih Bay rather than northern Maluku. There are as yet, however, no dated archaeological excavations from the relevant parts of western New Guinea and its offshore islands such as Biak and Japen. PEMP broke up into Proto Oceanic and Proto South Halmahera/West New Guinea.

an interaction that took tuber cultivation back into Island Southeast Asia? [...] Could neolithic I have extended from Borneo into Sulawesi, the Philippines and Maluku, and thereby have facilitated in some way the later and rapid expansion of neolithic II? To what extent does neolithic I represent a dispersal of farmers, as opposed to the movement of fragmented Neolithic technologies and ideas to resident island populations [...]?” (Anderston 2005:39). Claims for 5000 BP pottery in New Guinea have been discussed earlier in this paper. The undoubted spread of tuber cultivation from a New Guinea centre (sensu lato) into Island Southeast Asia does not require any necessary stimulus from that latter area, and seems to be a preceramic — and in that sense ‘pre-neolithic’ — phenomenon (Denham 2004; cf. Spriggs 1996a:335, 338). Anderson provides no evidence and I can find none either for a spread of ‘neolithic I’ beyond Borneo and Java into the Philippines and eastern Indonesia. At the local level, such as in parts of Sulawesi (Bulbeck et al. 2000), there undoubtedly was neolithic technology transfer (fragmentary or not), but what evidence we have for this in eastern Indonesia — with the exception of the spread of the domestic pig (see later discussion) — seems linked to the spread of ‘neolithic II’ populations speaking Austronesian languages (cf. Bellwood et al. 1998:262–270).
After the break-up of PCEMP as it spread into northern Maluku and/or parts of western New Guinea around 3500 BP, the languages in the CMP area diverged in situ, forming a linkage rather than a perfectly discrete subgroup.

### 4.6 Proto Oceanic (POc) and Proto South Halmahera/West New Guinea (PSHWNG)

POc was formed when a group of Eastern Malayo-Polynesian speakers moved along the north New Guinea coast to the Bismarck Archipelago (the New Britain, New Ireland, Mussau and Admiralty Islands area). This spread is associated with the beginning of the Lapita culture in the Bismarcks, now best dated to about 3300 BP (Specht and Gosden 1997). Kirch (2001) has argued for a somewhat earlier start based on five dates from the Mussau or St Matthias Islands, suggesting Lapita beginnings at 3550–3450 BP. The calibrated age ranges of all five dates continue post-3300 BP and this is really an argument around the margins unless it can be shown that the Mussau sites contain a distinctly earlier pottery style associated with these determinations.

POc is not very different from PMP, and according to one linguist’s best guess this would imply a period of perhaps 600 years between the two (Pawley 1999:125). A similar period prior to the earliest Lapita sites in the Bismarcks would bring us back to about 3900 BP, and this is indeed at the earliest possible age range of the first neolithic sites identified south of Taiwan and the Batanes. One might suggest that a separation more like 4–500 years between PMP and POc is a better estimate on the basis of the latest archaeological evidence. POc began to break up when speakers spread beyond the Bismarck Archipelago down the Solomons chain to Vanuatu and New Caledonia about 3150–3100 BP, crossing for the first time into Remote Oceania.

Although there were many resultant subgroups, for the purpose of this paper we will only go on to consider Proto Central Pacific, the ancestor of Fijian, Rotuman and all Polynesian languages, with a short discussion at the end concerning the Nuclear Micronesian languages and Yapese.

---

8 Space precludes dealing with the internal subgrouping of the Oceanic Austronesian languages of Near Oceania, and any associated archaeological evidence. The general view would associate the spread of pottery along the north coast of New Guinea from east back towards the west at about 1700–1500 BP with the spread of the North New Guinea Cluster of Western Oceanic languages along the same axis (for the linguistics see Ross (1988:Ch.5); for the archaeology see Lilley (1999, 2004) — but see also Specht et al. (2006); Terrell and Welsch (1997)).

Similarly, spread of the Papuan Tip Cluster from east to west along the south Papuan coast has been associated with the spread of Lapita-derived pottery-using cultures in the same area at about 2000 BP (Ross (1988:Ch.6, 1998); for the archaeology see Irwin (1991), Lilley (1999)). Most recently, however, pottery has been found as far west as the Torres Strait Islands (Carter et al. 2004; McNiven et al. 2006). Dates for this are thought to go back as far as 2600 BP. This has led to the suggestion that the spread of the Papuan Tip languages may have taken place along this coast up to 600 years earlier than has so far been attested in the archaeology. If this dating is sustained, it would be the only example in the entire region where the neolithic dates are in fact getting older rather than younger with further research!

Sheppard and Walter (2006) have recently considered the archaeology of the main Solomons chain in relation to linguistic boundaries there, particularly the ‘Tryon-Hackman line’ separating Meso-Melanesian languages of the Northwest Solomonic group from those of the Southeast Solomonic subgroup of Oceanic (Ross 1988:Ch.7). They make many interesting observations but their interpretations would seem to be in considerable conflict with current linguistic models. A linguistic evaluation of their paper would be most welcome.
PSHWNG diverged from Oceanic after about 3300 BP, representing the Eastern Malayo-Polynesian ‘stay at homes’ whose languages diverged in situ after this period.

4.7 Proto Central Pacific (PCP)

By 3000 BP there were Lapita sites in Fiji and Tonga, although stylistically the earliest Lapita pottery found in this region is from Fiji (Clark and Murray 2006). It is Fiji too where PCP is believed to have been spoken initially as a chain of dialects. Then settlement spread rapidly to Tonga and perhaps up to 200 years later to Samoa, although earlier Lapita sites may one day be found in this latter archipelago. PCP split up into an eastern form ancestral to Polynesian and some eastern Fijian dialects and a western form ancestral to Rotuman and some north-west Fijian dialects (see Green 1999:9 for further detail on the break up of PCP).

4.8 Proto Polynesian (PPn) and subsequent interstages

PPn is an extremely well-defined subgroup (see Figure 4), suggesting a long period for its development separate from the other Central Pacific languages to the west. Its distribution clearly included Tonga and Samoa, and probably some other nearby western Polynesian islands. Pawley (2004:255) suggests a period of common development lasting between about 1000 and 1300 years, albeit on somewhat shaky glottochronological calculations and comparison of core vocabulary and grammatical diversity with the Romance languages in western Europe. PPn broke up eventually into Proto Tongic and Proto Nuclear Polynesian (PNP), the latter spoken in the Samoan region. PNP split up when what became (East) Uvean and (East) Futunan started to diverge from what then became Proto Ellicean (PEc). The latter includes the languages of Samoa, Tuvalu and Tokelau. It is from the Samoan area that a group of speakers moved into eastern Polynesia, perhaps via Tuvalu and/or Tokelau.

4.9 Proto Eastern Polynesian (PEP)

The break-up of Proto Nuclear Polynesian has to be dated by the first appearance of people in areas to the east. As we have seen above, the dates for this are controversial. Very little archaeology has been done in either Tuvalu or Tokelau to evaluate the linguistic model for their primacy in settlement of eastern Polynesia. The Cook Islands would seem geographically to be the most likely place for early settlement. Yet they are small islands that may have been leapfrogged over in favour of settlement in Tahiti and/or the Marquesas. If we accept Fischer’s (2001; cf. Green and Weisler 2002) concept of a Southeastern Polynesian subgroup splitting first off the main tree, the PEP homeland could have been in the Tahitian area rather than in the Marquesas, although its successor stage Proto Central Eastern Polynesian (PCE) presumably was spoken in both archipelagoes.
Figure 4: Subgrouping of the Polynesian languages (adapted from Marck 2000 and Green and Weisler 2002)
There is currently no archaeological consensus on when eastern Polynesia was first settled. Spriggs and Anderson (1993) accepted dates in the range 300–600 AD for the Marquesas. Anderson (2002) now suggests Tahiti and the Marquesas were only settled around 900–1000 AD, and Pearthree and Di Piazza (2003) would push settlement of the whole of east Polynesia forward into the 1100–1400 AD bracket. However, if settlements of Hawaii and south-eastern Polynesia are acceptably dated in the interval 700–900 AD, then Pearthree and Di Piazza’s estimates are far too late. Anderson’s dates for French Polynesia also seem too late by at least a couple of centuries. I would not be surprised if settlements in the 500–800 AD range are one day to be found in either the Cook Islands and/or Tahiti and/or the Marquesas. My only difference with Green and Weisler (2002) on this is that they believe that the evidence has already been found there, whereas I do not.

4.10 Proto Southeastern (PSE) and Proto Central Eastern (PCE) Polynesian

The Southeastern languages consisted of Rapanui (Easter Island), original Mangarevan, (original) Eastern Tuamotuan and the now extinct languages of Pitcairn and Henderson Islands. There is no consensus figure for the ages of these interstages either. Green and Weisler push settlement of the Mangareva, Pitcairn and Henderson area back to 700 AD. There are extremely plausible dates from the region including Easter Island in the range 800–900 AD, but Hunt and Lipo have recently argued for settlement of Easter Island at about 1200 AD, while admitting it could go back to 1050–1150 AD. I would argue for 800 AD being around the right time for the break-up of the PSE subgroup, and similarly for the break-up of Proto Central Eastern (PCE), at a time when regular communication between the Marquesas and Tahiti may have ceased. PCE split into Proto Marquesic and Proto Tahitic.

4.11 Proto Marquesic (PMq) and Proto Tahitic (PTa)

PMq split up when a group of its speakers made the long journey north to discover the Hawaiian Islands, probably somewhere in the interval 800–900 AD. Southeastern Marquesan later came to influence Eastern Tuamotuan and Mangarevan, originally Southeastern Polynesian languages derived from PSE. This influence begins around 1150 AD or slightly earlier. Soon after this, at around 1200 AD, it was the Marquesan-influenced Mangarevan that was introduced to Rapa in the Australs, presumably by its initial inhabitants.

Proto Tahitic started to split up into its constituent languages of Western Tuamotuan, Tahitian, Rurutan, Common Austral, Cook Island Maori and New Zealand Maori some time prior to settlement of New Zealand in the period 1200–1300 AD, probably by the time the Cook Islands may have been first permanently settled at about 1000 AD. Settlement dates for Rurutu and the rest of the Australs (excluding Rapa), and for the Tuamotus have not yet been established, although there are now dates from the Peva site on Rurutu back to 1219–1288 AD (Bollt 2005). Later contacts between Tahiti and Hawaii had given a superficially Tahitian cast to Hawaiian language and culture by about 1400 AD.
4.12 The Micronesian languages

As already mentioned, the languages of the Mariana Islands and Palau are Western Malayo-Polynesian languages, resulting from two direct migrations from somewhere in the Philippines/Sulawesi region at about 3500 and 3400 BP respectively. The language of Yap is clearly a very early offshoot from Proto Oceanic. Although the earliest archaeological sites there only date to within the last 2000 years, it is likely that fully-fledged Lapita culture sites dating to about 3300 BP remain to be discovered. All the other Micronesian languages are derived from Proto Micronesian, a language derived post-Lapita from somewhere in the area between Manus and Central Vanuatu. Blust (1984) once favoured the south-east Solomons, particularly the island of Malaita. There is, however, as yet no archaeological record from that region of the requisite time period with which to compare the earliest assemblages of the Micronesian high islands of Truk, Pohnpei or Kosrae. These at present date to about 2000 BP, no earlier than some of the atolls in Micronesia, but occupation might be expected to go back somewhat further. How much further is as yet unknown.

The pottery found on the high islands is predominantly a plain ware, sometimes with lip notching. Comparable post-Lapita assemblages in northern and central Vanuatu date to the period 2800–2300 BP, at which time they are replaced by more highly decorated wares (Bedford 2006; author’s unpublished research). Lapita-derived plain wares of similar age are known from areas to the north-west in the western and northern Solomons and in parts of the Bismarck Archipelago, including Manus. Such assemblages — and similar ones may some day be recovered in the south-east Solomons — would provide suitable ancestors for the Micronesian pottery styles. If so, settlement of at least the high islands there must predate 2300 BP and could be several hundred years earlier.

5 Conclusions

This paper has sought to answer a query once raised with me by Andy Pawley concerning the tendency for archaeologists to argue for ever-younger dates for Austronesian spread within Island Southeast Asia and the Pacific region. I have attempted to explain why this has occurred by giving an historical review of the situation, using Pawley and Green (1973) as a convenient starting point. There is now something approaching a consensus in dating parts of the Austronesian spread, particularly in the region between the Bismarck Archipelago and west Polynesia. For other areas and interstages, however, there is either insufficient evidence to provide any dating framework, for instance in much of the more eastern part of Island Southeast Asia and for Tuvalu and Tokelau, or there are widely divergent interpretations of the data available, as in the whole of East Polynesia. Archaeology is now moving beyond an initial exploratory stage in many island groups. Over the next decade or so we can expect further elucidation of initial neolithic settlement dates. Given the error ranges of many radiocarbon determinations and the exigencies of the calibration curves to convert them to calendar years, however, the fine tuning needed to examine the correlation of archaeology and linguistics in some of the more recently settled parts of the Pacific may never be achieved.
I have already quoted out of context from Bob Dylan and the Beatles, doubtless sixties icons of the young Pawley, in this paper. I shall end by asserting, with Mick Jagger and Keith Richards, that for linguists at least, archaeological evidence may be a case where: ‘You can’t always get what you want, but [...] you get what you need’.

References


Emory, Kenneth P. and Yoshihiko Sinoto, 1969, *Age of the sites in the South Point area, Ka’u, Island of Hawaii*. Honolulu: Department of Anthropology, Bishop Museum.


Green, Roger C., 1988, Those mysterious mounds are for the birds. *Archaeology in New Zealand* 31:153–158.


‘I was so much older then, I’m younger than that now’


O’Connor, S. and P. Veth, 2005, Early Holocene shell fish hooks from Lene Hara Cave, East Timor, establish complex fishing technology was in use in Island Southeast Asia five thousand years before Austronesian settlement. *Antiquity* 79:249–256.


The impact of a dynamic environmental past on trade routes and language distributions in the lower-middle Sepik

PAMELA SWADLING

Introduction

Today small ships can travel up the Sepik River as far as Ambunti some 200 km as the crow flies from the sea (Figures 1–2). In pre-European times, contrary to expectation, trading activities did not extend from the coast up the navigable length of the Sepik River. Groups living in the middle Sepik, that is the stretch of river from above Angoram upstream to Ambunti, received coastal trade items not via the Sepik River, but overland from the north coast. Coastal trade goods were brought up the Sepik River but were diverted in its lower reaches up its southern tributaries. The latter in turn linked with routes into the central highlands. This paper examines how this cultural fence may have arisen.

The trade network that extended from the coast to the lower Sepik River and then diverged up its southern tributaries, as far as the flanks of the central highlands, is called in this paper the lower Sepik trade sphere (Figure 3). It was one of two trade spheres operating in the mid-lower Sepik basin in the early twentieth century. The other trade sphere, called here the middle Sepik trade sphere, extended from the north coast by foot over the coastal mountains and thence by foot and/or canoe to the middle Sepik. The southern links of this sphere, extending from the middle Sepik floodplain to the central highlands, were severely disrupted in about 1905 when the floodplain dwelling Chambri fled south to the foothills after being attacked by warriors from an Iatmul village (Gewertz 1983:105).

1 I wish to thank Laurie Bragge, Ulrike Claas, Terry Hays, Robin Hide, Paul Roscoe, Robin Torrence and Patricia Townsend for their comments on drafts of this paper and John Burton for sourcing the highlands axes from the Sepik-Ramu in the PNG National Museum. Any errors that may remain are my responsibility.
In addition to the long distance networks that linked coast, floodplain and highlands there were local trade networks that were primarily concerned with trading local products. It is proposed here that the long distance and local networks of the two trade spheres played a major part in producing the current language distributions found in the Sepik-Ramu region.

It is possible to gain some insights into the antiquity of coastal–highlands interaction in the Sepik-Ramu from prehistoric stone artifacts. These were made and used in the early to mid-Holocene. During this period an inland sea existed in the Sepik-Ramu region.

![General location map](#)

**Figure 1:** General location map

**The inland sea**

For much of the last 10,000 years the Sepik-Ramu floodplain did not exist. Instead there was a brackish inland sea that reached its fullest extent 6500 to 7500 years ago (Figure 4). This inland sea made it possible to voyage from the Bismarck Sea to the foothills of the central highlands. The Sepik and Ramu rivers and many smaller rivers drained into this inland sea via deltas (Chappell 2005, Swadling 1997). Freshwater originating from the central highlands and coastal mountains and tidal flows from the Bismarck Sea flowed through two passages located on either side of Bosmun Island situated near the mouth of the inland sea (Figure 4). The western passage was near Marienberg and the eastern near Bunapas. The presence of coral and white sand at Marienberg suggests that the circulation pattern within the inland sea operated with seawater entering via the Marienberg passage and brackish water discharging by the Bunapas passage. This would explain why mid-Holocene brackish water shellfish middens (Swadling 1997) occur in the vicinity of the Bunapas passage (Chappell 1993, 2005).
The impact of a dynamic environmental past on trade routes and distributions

Different styles of stone mortars and pestles and their distributions during the time of the inland sea provide an indication of interaction during the mid-Holocene (Figures 4–5). Pedestal mortars, for example, demonstrate that there were links between the highlands and shores of the inland sea, then east along the New Guinea coast as far as Oro (Swadling, Wiessner and Tumu 2008). Other finds shown in Figure 5 show that these links extended from the highlands along the coast to the Huon Peninsula across Vitiaz Strait to West New Britain, as far as the Willuamez Peninsula (Figure 1). Independent confirmation that links existed between West New Britain and the lower Sepik region in the mid-

Figure 2: Location map
Holocene is provided by obsidian artifact finds. Four stemmed obsidian tools have now been found in the lower Sepik region. One is unavailable for study, but the remaining three have been sourced to the Willaumez Peninsula (Swadling and Hide 2005:307–308). They were manufactured in West New Britain from some time before 6160–5740 years ago and ceased to be used by 3480–3160 years ago (Torrence and Swadling 2008).

Upstream of both the Sepik and Ramu deltas are large clusters of stone mortar and pestle finds (Figure 4) suggesting that these areas had higher populations than surrounding areas in the mid-Holocene. Such artifacts have been dated in the highlands from 8000 to 3000 years ago (Swadling and Hide 2005:293). The finds from the vicinity of Kambot may have come from small islands in the former inland sea. Some are shown in Figure 4, but spot heights on the floodplain indicate that there would have been many other islands in the inland sea. Other mortar and pestle finds have been reported from the north coast and offshore islands such as Kairiru, Muschu and Manam.

The presence of a mortar as well as a pestle in the foothill country drained by a tributary of the Karawari, and a stone figure from the upper reaches of another Karawari tributary are significant as these finds lie on an important trade route between the floodplain and lower Lagaip valley in the highlands (Swadling, Wiessner and Tumu 2008). Another indication that this trade route dates back to the mid-Holocene are the trochus armbands\(^2\) that were excavated from deposits of that age in the Kutepa rock shelter located in the

\(^2\) Trochus armbands were not observed in the highlands in the 1930s (Hughes 1977:185). It is also relevant to note that few *Trochus niloticus* occur on the modern Sepik coast, but the species is common in the Vitiaz Strait region. The presence of stemmed obsidian artifacts of mid-Holocene age from West New Britain in the lower Sepik makes it likely that the trochus armbands found in the Kutepa rock shelter originated from the shores of Vitiaz Strait.
lower Lagaip (pers. comm. Jo Mangi 1988). It is on this same trade route at an altitude of 600 metres that the Yafebaiyeli sago swamp is located. This ten square acre sago swamp is located in an east-west fold in the northern flanks of the central highlands (Dornstreich 1974:140–141, Map 4) and was possibly established as a source of food for trading parties. Yen (1985:837) considers sago above 300 metres to have been planted. The altitude, size and location of this swamp on this trade route make it a prime site for a palaeoenvironmental study.

Apart from the above-mentioned route, there were other routes via different Karawari tributaries into the highlands from the shores of the inland sea. East of the Karawari there were also routes via the upper Yuat and its tributaries (Gorecki and Gillieson 1989), the upper Keram (Flanagan 1983:13) and rivers that drain into the Ramu (Görlich 1998; Kaspruś 1973; Majnep and Bulmer 1977).

Figure 4: The Sepik-Ramu inland sea at its fullest extent 6500–7500 years ago. Also shown in the distribution and frequency of prehistoric stone mortars and pestles (shoreline follows Chappell 2005).
The formation of the floodplain and emergence of two interaction spheres

After 4000 years ago the inland sea started to infill, resulting in the formation of the current floodplain and the seawards progradation of the coastline (Chappell 2005). When the Sepik-Ramu floodplain formed, its drainage pattern was different to what exists today. The key difference was that the Ramu River did not follow its current direct course to the Bismarck Sea, but was actually a tributary of the Sepik. Formerly the Ramu River flowed west in the vicinity of Annanberg and then followed the current course of the Keram River (Figure 4). Aerial photos such as those available on Google Earth clearly show that the channel of the Keram River is superimposed on larger meander bends than the Keram’s flow capacity can produce.\(^3\) The Ramu has changed its channel position more than once as another of its abandoned palaeochannels can be seen between the current channels of the Keram and the Ramu (Yu et al. 1991). River channel changes such as that of the Ramu have implications for social interaction on and across the floodplain, as river transport is the only means of travelling any distance in this environment.

After 4000 years ago access to the Sepik basin from the coast was only possible via two routes. These routes were via the Sepik River or overland from the north coast. Control of the Sepik River would have become a strategic concern. It is likely the Ramu and Sepik groups who produced the clusters of mortars and pestles found upstream of the Ramu and Sepik deltas would have competed to gain control of the main Sepik channel. The Ramu group were able to divert trade from the lower Sepik River to its southern tributaries, especially the Ramu/Keram, Yuat and eastern Karawari rivers. The latter was then probably flowing via a palaeochannel that is now Kenglame passage. Such a scenario would explain the origins of the lower Sepik trade sphere.

A very different scenario explains the origins of the middle Sepik trade sphere. The descendants of the Sepik groups would have found that they were cut off from supplies of exotic north coast trade goods when trade items could not be obtained via the Sepik River. This would have encouraged them to move away from the developing floodplain and back swamps and to try to re-establish their coastal links via the coastal mountains to the north. This need to relocate and regain former links may explain the current distribution of speakers of the Ndu language family (Figure 8). People speaking languages from this family are found mainly north of the main river in the middle Sepik. Their distribution extends into the foothills of the coastal mountains and in the case of the Boiken speakers north to the coast and offshore islands (Allen 2005; Roscoe 1989, 1994; Swadling and Hide 2005). Like the speakers of the long-present Torricelli family (Laycock 1973), the ancestral speakers of Ndu languages established trade networks that extended from the north coast to the middle Sepik.

The long distance trade networks

Historical descriptions of trading activity are available for the main environments within the lower Sepik trade sphere, but this is not the case for all environments in the middle Sepik trade sphere. The least known section of the latter concerns the route from the southern part of the floodplain up the northern flanks into the highlands, but Kumagai (1998:49) provides some information on links across the floodplain to the foothills. This

---

\(^3\) Archaeologists may find that surveying the banks of former palaeochannels, especially the Keram, may turn out to be a productive way of finding prehistoric settlement sites.
The impact of a dynamic environmental past on trade routes and distributions

lack of observations can probably be explained by the major disruption to life on the middle Sepik floodplain that occurred early in the twentieth century. This was the dispersal of the Chambri in about 1905 after attacks by Iatmul from Parambei village. The Chambri did not return to their island for more than 20 years (Gewertz 1977, 1983:105–106, 220). However, the wide swathe of people speaking Sepik languages, south of the middle Sepik River and up onto the northern flanks of the central highlands, and in the case of the Hewa into a highlands valley (Lagaip), suggests that trade networks once extended south of the middle Sepik into the highlands (Figure 3).

Long-distance traders

Historical accounts can provide us with some indication as to the nature of trade in different environmental zones in the prehistoric past. Coast to highlands trade is dependent on chains of interpersonal relations with the nature of the environment influencing the actual distance travelled in any particular segment. Within the social networks that make up the lower Sepik interaction sphere two segments stand out as being longer than others and thus critical for maintaining links between the coast and the highlands. These segments cover from the coast up the lower Sepik, on the one hand, and from the edge of the floodplain to the edge of the highlands, on the other. At the beginning of the twentieth century the Murik dominated both the coastal and lower Sepik trade (Lipset 1985, 1997; Tiesler 1969–70), whereas many different groups were involved in moving goods between the floodplain and the highlands. The Gadio who live on the upper Karawari and its major tributary the Korosameri provide the basis for the observations given here.

In the first half of the twentieth century the Murik dominated riverine trade up the lower Sepik as far as their ancestral homeland at Moim Lagoon. They were also the leading seagoing traders both east and west of the Sepik delta. Their trading partners extended westwards along the coast as far as Yakamul in West Sepik province, on the offshore

4 This major disruption arose because the political economy of the middle Sepik changed following the importation of iron tools. These were brought into the middle Sepik from the north coast via the traditional trade sphere. In the late nineteenth century a number of trade stations were established on the Sepik coast. They were manned by either German or Ternate traders (Swadling 1996:219–220). By about 1905 sufficient iron tools had been imported into the middle Sepik to make the stone tools produced by the Chambri redundant. Accordingly the Iatmul from Parambei village no longer felt any need to protect their source of stone tools and raided the Chambri, who were forced to flee. The Parambei attackers even had a shotgun acquired from the Germans (Gewertz 1977:38–39, 1983:105).

5 According to their oral history the Murik moved to the coast 200 or more years ago (Somare 1975:15; Lipset 1997:28). It is not clear what prompted their move from just upstream of Angoram to the coast. Perhaps visits by European ships including the visit in July 1616 by the Dutch sailing vessel commanded by Jacob le Maire might have been a significant factor. The journal of the voyage by Jacob le Maire and his captain Willem C. Schouten reports passing Manam and then sailing through what was assumed to be the discharge of rivers. The sea in the latter area was observed to have reduced salinity, to be coloured by sediments and to have trees, branches and leaves floating on its surface. This description fits the nature of the discharge of the Sepik and the Ramu rivers. Le Maire and his crew anchored at two locations just west of the Sepik River mouth for three days. During this time they received two visits from the inhabitants of the East Sepik coast. On the last day of their visit twenty canoes containing, men, women and children came along side. The Dutch were disappointed in not being able to obtain one of the pigs they saw in the nearby villages. Coconuts were scarce and not cheap, the ship’s journal records that the locals wanted a fathom of linen for four nuts (De Villiers 1967:219–221). The inclusion of women and children in the canoes going out to the ship suggests that this was not the first time that a ship’s crew had traded with the people in this part of New Guinea.
islands from Walis to Manam and eastwards along the coast to just beyond the mouth of the Ramu River. Despite the presence of Austronesian speaking communities (Figures 2 and 6) along the coast and offshore islands, the Murik dominated in both the maritime and riverine movement of goods. This raises the question as to whether during the time of the inland sea the population clusters living upstream of the Sepik and Ramu deltas had a similar hegemony over their respective regions.

The other long segment within the lower Sepik sphere involved carrying up goods from the floodplain and bringing down items from the highlands. The Gadio described here played a comparable role to that of the Mountain Arapesh who moved goods over the coastal mountains (Mead 1967:22; Dobrin and Bashkow 2006). The people living on the northern flanks of the central highlands are renowned for the distance they travelled on foot. In some cases they would take three to five days to reach trade partners (Hatanaka 1988). The Gadio were one of the groups involved in such activity in the headwater area of the Korosameri and Karawari rivers (Figure 3).

In 1967–68 taro was the Gadio’s staple. They used sago when garden food was in short supply and as provisions when travelling. Yafebaiye, mentioned above, was their main sago swamp. The Gadio’s most important trade route was the one that linked the highlands and the Sepik lowlands. This interest influenced marriage patterns amongst this predominantly patrilocal group. To the north they obtained brides from the Alamblak living at the edge of the floodplain and to the south from the Nete living on the highlands fringe. The Alamblak depend on sago, collecting, hunting and fishing for their livelihood and had haus tambaran, religious sculpture and male initiation ceremonies, whereas the Nete depend on sweet potato cultivation and pig husbandry. In the 1960s the Gadio regularly visited their affinal kin and co-resided with them during their visits, shared economic work and exchanged goods (Dornstreich 1974:37, 122, 198, 216–217, 494, 504–505, 518; Foley 1991:6, 10). Overall the Gadio’s main contribution to the trade network was labour, their efforts moved goods from the foothills to the highlands fringe and vice versa.

The north coast to highlands chain of interpersonal relations did not end in the highlands but continued south across New Guinea. For instance, the lower Sepik route into the highlands via the Gadio, described above, continued south from the lower Lagaip to the Tari basin and thence south to the Papuan Plateau, where links extended to the south coast. As with other groups located on this chain the Tari considered this alignment across ecological zones to be their most important trade route (Ballard 1994:141).

**Exotic trade items**

Exotic items from the coast that reached the highlands included shell valuables, such as shell rings and Nassa shells. Dances and ceremonies were also traded upriver from the coast. For instance, Mead (1948:172) reports that the Biwat of the middle Yuat occasionally obtained these from a lower Yuat village in the early twentieth century. The antiquity of the trade in dances and ceremonies is an interesting question, as it has implications for interpreting prehistoric ceremonial activities such as the wide distribution of bossed stone mortars in eastern New Guinea as well as their distribution east as far as the Willaumez Peninsula of West New Britain.

---

6 The desire to learn the latest dance seems to have provided a major incentive for trade in some areas. For example, the Beach Arapesh recognised that the Murik used the selling of dances as a way of promoting their shell valuables, baskets and other products (Mead 1970:35).
From the highlands and its flanks came a different range of exotic items. In 1932 Mead, based at Biwat on the middle Yuat, recorded that bird of paradise plumes, pearl shell crescents, hunting magic, stone axes, stone sago cutters, bows and arrows were traded from the highlands to the Sepik floodplain (Mead 1948; McDowell 1989, 1991). How far bird of paradise plumes from the highlands were traded downriver is not documented. However, some indication of the importance of highlands trade goods in the Sepik-Ramu basin is evident in the distribution of stone axes and pearl shell breast ornaments on the floodplain.

Figure 5: In the mid-Holocene long-distance links between the highlands, coast and West New Britain were made possible by the presence of the inland sea. The comparable mortars and pestles found in these regions are the legacy of this contact.
The distribution of stone axes from the highlands on the Sepik-Ramu floodplain, apart from an odd leakage, falls within the area of the lower Sepik trade sphere. Plotted on Figure 8 is the recorded distribution on the floodplain of stone axes from the Jimi and Wahgi quarries in the highlands as well as the extent of the Chambri stone tool trade. The middle Sepik floodplain was supplied by a local source at Chambri Hill (Gewertz 1977:87–90, Pétrequin and Pétrequin 2006:309–316). The inner ring represents the direct trade by the Chambri with their trade partners, whereas the outer ring shows the onward trade made by these trade partners. It is not known when production at the Chambri quarries commenced, but they ceased production when the Chambri abandoned their island in about 1905 (Gewertz 1977:38–39, 1983:105). By this time stone tool use was in decline throughout the lower-middle Sepik floodplain.

The pearl shell crescents being traded down the Yuat in 1932 seem to have been more important to the people of the middle Sepik trade sphere than those of the lower Sepik. In the middle Sepik they are the second most important shell valuable for the Iatmul and are third for the Chambri (Milan Stanek pers. comm. 1985; Gewertz 1983:238). The German expeditions up the Sepik made respectively in 1909 and 1912–13 report the presence of pearl shell crescents at Malu near Ambunti and at a Iatmul village, Parambei No. 2 (Reche 1913; Kelm 1966: Plate 125). They were not common.

Neither Mead nor Gewertz make specific comments as to how the Chambri received their pearl shell valuables. For Gewertz (1983:77) shell valuables originated from the north and reached the Chambri through the Iatmul and left the region to the south through Sepik Hills groups. As the Torres Strait is the source of the pearl shell used to make the crescents, it is likely that the Chambri either received them from groups living near the Korosameri or Karawari Rivers or indirectly from these groups via the Iatmul. Here it is relevant to note that pearl shell crescents feature in the prehistoric rock art of the Karawari found in rock shelters on the flanks of the central highlands (Swadling et al. 1988:19, Plate 45). This area is close to where the best quality and greatest number of pearl shells were observed in the highlands in 1933. This was the upper Wahgi of the Western Highlands and the first Europeans in the area were told that the pearl shell had been obtained from the inhabitants of the Kaigel and Nebilyer valleys to the south (Hughes 1977:55, 57, 185). Pearl shell artifacts have been traded across New Guinea for a long time, as broken pieces of pearl shell dating from 3000 to 2000 years ago have been excavated from the Ritamauda rock shelter in the Yuat gorge (Swadling and Anamiato 1989:225–227).

The local trade networks

Historical accounts indicate that local networks extended out from nodes along the long distance networks. For instance, the Biwat of the middle Yuat were not only a node in the long distance network, but were also the focus for the local trade network that extended east and west across the middle Yuat floodplain. Some exotic items were traded in the local network, but most items were local products such as pottery, spears, paddles, net

---

7 Mead observed during the four months she spent with the Chambri in 1933 that one kina shell changed hands four times. Some idea of the importance of this valuable is gained from the nature of these transactions. They were as follows: Tchuikumban received the kina shell from his classificatory sister during a ceremony to open a men’s house in return for his contribution of sago and coconuts; he gave it to his prospective father-in-law’s clan to help them fulfil their debt to his fiancée’s dead husband’s father; but took it back when he feared his wedding was off; he then contributed it to a pig payment his clan made to their affines in order to end a major mourning ceremony (Gewertz 1983:75).

On the flanks of the central highlands ceremonies provided an opportunity for trade between groups living in the same environmental zone. Dornstreich observed for the Gadio that aggression and ritual were closely related, but mutually exclusive events. A *yabo* ceremony would be held to temporarily end the prevailing sporadic raiding and this would provide the opportunity for trade. For the Gadio such occasions allowed them to obtain goods such as arrows and formerly stone axe blades. Such ceremonies were in decline by the late 1960s (Dornstreich 1974:508–509, 513, 516).

**Language families, interaction spheres and trade networks within the lower-middle Sepik**

Environmental changes that date back at least 4000 years have influenced the development of the language families found in the lower-middle Sepik. It is likely that the formation of the floodplain played a part in their distribution pattern and that protoforms were spoken on the shores of the inland sea during the mid-Holocene. This proposed antiquity is comparable to that proposed for the language groups in the Eastern Highlands on the one hand and the Wahgi communities in Simbu and the Western Highlands on the other (Foley 1986; Swadling 2005:11).

![Modern language map](image)

**Figure 6:** Modern language map  
(Sources: Laycock 1981; Dornstreich 1974; Foley 1991, 2005; Ross 2005)
Looking at the distribution of language families (Figure 6) as well as interaction spheres and the trade networks within the latter (Figure 3), it is apparent that there is a close relationship between language and trade in the lower-middle Sepik and Ramu. Trade was undertaken by using canoes, as these were the only way of travelling any distance across the floodplain. Within the lower Sepik trade sphere related languages occur along the trade networks that developed along each major waterway. This explains why the Lower Sepik, Yuat and Ramu language families have a generally linear distribution.

Trade networks within the other proposed trade sphere, which extended from the north coast to the middle Sepik (Figure 3), brought coastal goods, such as shell valuables, to the middle Sepik. The absence of settlements until relatively recently along the Sepik River, from upriver of Angoram to downriver of Ambunti\(^8\) may explain how Lower Sepik language family speakers came to be resident at Chambri and formerly at Aibom.\(^9\) The Chambri and Aibom hills in the swamp country south of the Sepik River have stone and clay resources respectively. Although the Chambri stone trade prior to its demise about 1905 attracted large numbers of shell valuables (Gewertz 1983:238), little is known as to the nature of the trade networks that extended south of the middle Sepik into the highlands. The large number of Sepik languages spoken by people living along the foothills, northern flanks, and in the case of the Hewa within the highlands, suggests that active trade networks once extended up the western Karawari tributaries, as well as up the April and Wogamush rivers. Some of these links may date from the time of the inland sea and others may be more recent. Despite the possibility of varying time scales, trade networks provide a plausible explanation as to how languages within the Sepik language family came to be spoken from the north coast south into the highlands.

**Conclusions**

The long-term view presented in this paper shows that environmental changes can have significant impacts on the formation of trading relationships and spatial networks and on the distribution of language groups.

In the early to mid-Holocene a large inland sea covered much of the middle and lower Sepik-Ramu basin. This body of water allowed canoe access from the Bismarck Sea to the base of the central highlands. Ease of interaction between coast, lowlands and highlands was far greater across the inland sea than subsequently when the position of the rivers and tributaries of the floodplain restricted movement and changed access to former trade partners and resources. During the time of the inland sea, the people inhabiting the Sepik-Ramu had links not only with the highlands but also along the New Guinea coast and across Vitiaz Strait to West New Britain. This interaction is demonstrated by the presence in the Sepik of obsidian artifacts from West New Britain and the occurrence of particular styles of mortars and pestles in both east New Guinea and West New Britain.

---

\(^8\) The first Iatmul were Sawos who left the swamp forest north of the Sepik River and settled on the banks of the river. Iatmul genealogies indicate that this happened more than 250 years ago. Unlike the Kwoma who claim to have expelled or absorbed the previous occupants of the Ambunti hills, Iatmul traditions indicate that the banks of the middle Sepik River were unoccupied when they settled there (Newton 1997:379). By the early twentieth century Iatmul villagers dominated the middle Sepik floodplain.

\(^9\) The people at Aibom came under Iatmul influence post 1850 (Schuster 1990).
The impact of a dynamic environmental past on trade routes and distributions

Figure 7: Modern extent of the Ndu language family (following Layock 1981)
Note that *latmul* from Kandingai only settled south of Lake Chambri in 1944 at the invitation of a Chambri *luluai* (Gewertz 1983:138–140).

Figure 8: Distribution of highlands axes and extent of the Chambri stone tool trade
The inner dashed line indicates the extent of trade by Chambri with their own trade partners and the outer ring shows the onward trade by these trade partners (based on Gewertz 1977:Map 3.1).
All the rivers that formerly flowed into the inland sea, apart from the Sepik, the Korosameri and current channel of the Karawari, join the Sepik River in or near its lower section (Figures 2 and 4). Once the inland sea infilled, the available routes across the floodplain influenced the way the interaction spheres developed. The only means of travelling any distance across the forming floodplain was by river transport. This meant access routes between the coast and the highlands followed the pattern of the river channels on the floodplain. This pattern gave rise to the lower Sepik trade sphere and explains why in the early twentieth century the main flow of upriver trade from the Sepik delta was not up the main channel to the middle Sepik, but south from the lower Sepik up its southern tributaries, principally the Keram, with the Yuat and eastern Karawari system being also important trade routes. It is proposed in this paper that these trade networks played a major part in producing the language family distributions found in this part of the Sepik-Ramu floodplain.

A different situation seems to have developed for those groups living in the vicinity of the former Sepik delta of the inland sea. Once the infilling of the former inland sea began groups moved away from the developing floodplain seeking to regain coastal links via the coastal mountains to the north. This would explain the distribution of speakers of languages belonging to the Ndu family within the Sepik language family.

In summary, interpersonal trade relations within the middle-lower Sepik-Ramu can be grouped into two trade spheres. One trade sphere extended from the north coast through to the middle Sepik, the other from the Sepik delta to the southern Sepik tributaries, as far west as the Karawari. Speakers of languages belonging to both the Sepik family and long-present Torricelli family (Laycock 1973) participated in the coast to middle Sepik trade sphere, whereas the lower Sepik to southern Sepik tributaries trade sphere was operated by speakers of the Lower Sepik, Yuat and Ramu language families. These observations link linguistic distance and interaction spheres, and the trade networks within them, and raise interesting questions for both linguistics and archaeology.

References


De Villiers, J.A.J., ed., 1967, The East and West Indian mirror, being an account of Joris van Speilbergen’s voyage around the world (1614–1617), and the Australian navigations of Jacob le Maire. Nendeln: Kraus Reprint.


The impact of a dynamic environmental past on trade routes and distributions


~ Reconstructing language history ~
14  The amalgamation of Malagasy

ALEXANDER ADELAAR

1 Introduction

Malagasy is an Austronesian language spoken on Madagascar, off the south-eastern coast of Africa (see Map 1). Malagasy's rich morphology is in many ways more conservative and reminiscent of the structure of Proto Austronesian than the morphology of its south-east Barito sister languages in South East Borneo. These sister languages are spoken east of the Barito River in the northern part of South Kalimantan Province and in directly neighbouring areas in Central Kalimantan (see Map 2). As might be expected, not all present-day Malagasy morphology is inherited from Proto Austronesian, or even from Malagasy's more direct ancestral language, Proto South East Barito. Some morphological elements are due to recent grammaticalisation or language contact (or both). Dahl (1954, 1988) discussed some Malagasy affixes that he considered to be due to Bantu influence.

In the following pages I will discuss Dahl's evidence for grammatical borrowing from Bantu, and I will treat various morphological elements that seem to be the result of borrowing from Austronesian outside the South East Barito subgroup or of recent grammaticalisation of inherited lexical elements. This will provide a better understanding of the grammatical history of Malagasy and show the importance of language contact in the changes that have lead to the development of present-day Malagasy morphology.

---

1 This contribution was written with support from the Bijzonder Onderzoeksfonds (Fellowship SF/08/009) at Leuven University. I am especially grateful to Dr Jean-Christophe Verstraete and the Department of Linguistics (Erasmushuis) for their hospitality. I would like to thank Dr Philippe Beaujard (CNRS, Paris), Prof. Noel Gueunier (University of Strasbourg), Prof. John Wolff (Emeritus Prof., Cornell University) and the editors of this volume for their valuable feedback on an earlier draft of this paper, and to Mrs Chandra Jayasuriya for preparing maps. They are in no way responsible for any errors in the present version.

2 Maanyan is the only one for which we have reliable grammatical information (cf. Dahl 1951; Gudai 1985).
Map 1: Madagascar, Indonesia and the Indian Ocean
Map 2: The languages of South Borneo that are mentioned in this chapter
Malagasy has many dialects (see Map 3 for the location of those mentioned in this chapter). The island roughly consists of two main dialect areas. The ‘eastern’ area covers the east and centre (as well as a small section of the north-west). Its dialects are characterised by the change of the historical sequences *li and *ti to respectively di and tsi. They include Merina, the main dialect, which is spoken in the central highlands surrounding the capital Tananarivo and has provided the basis for standard Malagasy. The ‘western’ area covers the west, the south, and the northern tip, with Sakalava as its most notable dialect. Dialects in this area typically lost their historical final nasals (Gueunier 1986:148).

In an attempt to represent Malagasy words in this chapter in a more phonemic way, I have made the following adaptations to the official Malagasy spelling: I write ‘o’ as ‘u’, ‘y’ as ‘i’, and the final whispered ‘a’ as ‘á’ (in the endings -ná, -trá, and -ká). I furthermore write Sakalava ‘ng’ as ‘ŋ’.

Map 3: Dialects of Malagasy mentioned in this chapter

2 Dahl’s evidence for Bantu influence on Malagasy morphosyntax

In 1954 Dahl published an article on the influence of Bantu languages on Malagasy. In order to understand the interpretations he gave to the data that he identified as Bantu borrowings, a few explanatory words are in order about his underlying assumptions regarding the prehistoric relations between the early Austronesian migrants to east Africa and the Africans they encountered.

3 Note that in Madagascar the usual dialect names are somewhat imprecise, as they are based on ethnic and geopolitical criteria, rather than on linguistic ones. For instance, the Sakalava, Betsimisaraka and Bara groups each speak various dialects that do not necessarily belong to the same main dialectal divisions.
Following another Norwegian scholar of almost the same name (L. Dahle 1876–77, 1883), Dahl held the theory that these migrants had travelled to east Africa on their own accord, and that they had settled in Madagascar upon arrival. There they met an already existing Bantu-speaking population, which ended up adopting their language. This language shift brought about a ‘Bantu substratum’ in what would eventually become the Malagasy language. Dahl (1954:329) believed that the language of this early Bantu population was closely related to the Comorian languages, if it was not itself a Comorian dialect.

However, there is a general consensus among scholars that Madagascar had no Bantu population prior to the arrival of Austronesian speakers. Archaeologists used to believe that the island was uninhabited at that stage (Dewar and Wright 1993; Wright and Rakotoarisoa 2003). More recently, Roger Blench has argued that the occurrence of cutmarks on bones of extinct animals and vegetational change in the Malagasy landscape strongly suggest the presence of an earlier hunter-gatherer population from 400–300 BC onwards. However, the people in question were not Bantus. They may have been the precursors of the Mikea, a group of hunter-gatherers in South Madagascar who are culturally different from other Malagasy, although nowadays they speak a form of Malagasy (Blench 2007). Another point is that there is strong multidisciplinary evidence to suggest that the early South East Barito migrants first arrived on the east-African mainland, where they intermarried with local Bantu speakers before they finally settled on Madagascar as an Austronesian-speaking population of mixed Asian-African descent (Adelaar 2009). As I will show in the following pages, these new insights allow for a different interpretation of the history of Malagasy as well as of some of the affixes treated in this chapter. Dahl’s notion of a Bantu ‘substratum’ is somewhat premature, as there is no evidence of an early Bantu-speaking population in Madagascar, which adopted the language of the Austronesian newcomers. It is obvious that at some stage there have been Bantu speakers who shifted to that language, but it remains obscure where this happened, and in what sociolinguistic context. Considering the impact of Bantu languages, one could just as well think of an adstratum, which is more in line with the fact that in many ways, Malagasy has remained largely Austronesian. Only its phonology has undergone many changes, part of which are shared with Comorian languages (Dahl 1954:343–352).

Bantu influence in Malagasy probably comes from various sources, although there is a problem in identifying these sources with great exactitude. Dahl believed that they were the Comorian languages and Swahili, which all belong to the same Sabaki (coastal) subgroup of Bantu (Nurse and Hinnebusch 1993). He conceded that there were various lexical items that looked as if they originated from mainland Bantu languages spoken in the vicinity of Lake Victoria, but he discounted these as old words that had already disappeared in Comorian and Swahili (Dahl 1954:329). Wherever these words came from, in general it is evident that both Comorian and Swahili were among the languages that lent Bantu vocabulary to Malagasy. What remains uncertain is whether there were also other Bantu source languages, and furthermore, what the exact source language was of each individual Bantu loanword in Malagasy.

Comorian is a group of four very closely related languages (viz. Shingazija, Shindzuani, Shimwali and Shimaore), which in turn are rather closely related to Swahili. The latter is spoken on the African mainland, but also used to be a literary language on the Comoros until the early 20th century. Another Bantu language playing a role in the linguistic history of Madagascar is Makhua, because many (Makhua-speaking and other) slaves were
Alexander Adelaar

imported from Mozambique into Madagascar in especially the mid-19th century. However, Makhua apparently never became a major source of lexical borrowing to the Malagasy speech community.4

Dahl attributed three morphosyntactic phenomena to Bantu influence: (1) the nominalising prefix ki-, (2) the agent prefixes mpaN- and mpi-, and (3) the development of future tense. In the following subsections I give an assessment of his evidence.

2.1 Nominalising ki-

The nominalising ki- prefix is no doubt the most visible evidence of Bantu influence in Malagasy morphology. It occurs with Bantu roots as well as with roots of Indonesian provenance. Dahl (1954:352–353; 1988:120) noted that it is not productive in present-day Malagasy. He also indicated that it reflects the nominal prefix *ki- in Bantu languages and has meanings similar to Swahili ki-, which are as follows:

1. Prefixed to nouns, it forms diminutives, e.g. tranu ‘house’ and Merina ki-tranu-tranu, Sakalava ki-tranu ‘a miniature house for children’s game [sic]’ (< Malay5 dayaw ‘field hut’); lulu ‘ghost’ and Sakalava ki-lúlúke ‘butterfly’ (< Bantu, cf. Swahili ki-lulu ‘small insect’; Dahl 1988:109).

2. Prefixed to verbal roots it often forms instrumental nouns,6 e.g. fafa ‘to sweep’ and ki-fafa ‘broom’; lalau ‘to play’ and ki-lalau ‘plaything’.

3. In other cases where it is prefixed to verbal roots, it has a somewhat wider meaning,7 e.g. sese ‘to bring far away’ and ki-sese ‘competition in throwing far away’; ki-lalau, which also means ‘a play, game’.

Interestingly, Dahl did not note the occurrence of a variant tsi-. It has the same meaning as ki- and is sometimes in free alternation with it, e.g. the above example ki-tranu-tranu also occurs as tsi-tranu-tranu (same meaning). It must reflect a palatalised variant of ki- in the donor language. This variant is widespread among Bantu languages and sometimes occurs alongside ki-, as in Swahili, where the distribution of ki- and tši- is phonologically determined. In the Comorian languages, historical *ki- first became *ši- before it became hi- in Shingazija and š- in Shindzuani (Nurse and Hinnebusch 1993:197). It remains unclear whether Malagasy has both variants because of variation in the donor language or because of the existence of different donor languages.8

Historical initial and intervocalic *k as a rule became h in Malagasy, including in Bantu loanwords. However, *ki- remained ki-, a fact which Dahl (1988:121) attributed to a possible palatal pronunciation that this prefix might have had in the Bantu donor language. However, the change from *k to h happened early on in the post-migratory history of Malagasy, and while Bantu influence in Malagasy may have begun at a very early stage in that history, it must have continued until very recently through regular trade contacts with the Comoros and the African coast (including the slave trade). Therefore, a more likely

---

4 Prof. Noel Gueunier (email communication).
5 In this chapter, this term refers to various forms of literary Malay, including the Malay used in traditional literature, standard Indonesian and standard Malaysian Malay.
6 Dahl’s formulation: ‘an object somehow related to the action expressed by the root’ (Dahl 1988:120).
7 Dahl’s qualification as an ‘abstract noun’ is not entirely accurate.
8 It is likely that Malagasy borrowed from both Comorian languages and Swahili. Swahili influence is not surprising given its role as a literary language on the Comoros.
The amalgamation of Malagasy

explanation for the fact that *ki- did not become *hi- is that it is a relatively recent prefix: it remained ki- because it was borrowed into Malagasy after the *k > h change had ceased to be active.

2.2 The future tense marker /hu < h-/  
Malagasy verbs exhibit a threefold tense distinction. The present and past tenses reflect an older (Austronesian) aspectual distinction, but the future tense does not. The latter is expressed by the auxiliary hu, or by its shortened prefixed form h-, which often occurs before an initial vowel or replaces m- before verbs with a prefix beginning with m- in the present tense. Parallel to this, the past tense is usually marked with n- and nu- in the same positions where the future is marked with h- and hu-. For example:

```
ume-nə azi ni vula ‘the money is given to him’ (give-PASS 3s.OBL ART money)
numenə azi ni vula ‘the money was given to him’ (n-ume-nə PST-give-PASS)
humenə azi ni vula ‘the money will be given to him’ (h-ume-nə FUT-give-PASS)
mangalatrə Pauli ‘Paul steals’
nangalatrə Pauli ‘Paul stole’
hangalatrə Pauli ‘Paul will steal’
miakatrə ahu ‘I lift’
niakatrə ahu ‘I lifted’
hiakatrə ahu ‘I will lift’
```

Underived verbs do not distinguish between present and past tense; they are preceded by hu in future tense. For example:

```
tunga izau izi ‘she’s arriving now’ (tunga ‘to arrive’; izau ‘now’; izi 3s)
tunga umali izi ‘she arrived yesterday’ (umali ‘yesterday’)
hu tunga rahampitsu izi ‘she’ll arrive tomorrow’ (rahampitsu ‘tomorrow’)
```

Dahl traced the development of this three-way tense distinction to Bantu influence, which seems to be a reasonable analysis. In Sabaki Bantu languages, future is expressed by using the verb *caka ‘to want’ as an auxiliary before infinitive verbs. Infinitive verbs are preceded by *ku, which is basically a nominal prefix. In Swahili, *caka became taka ‘to want’ and was furthermore grammaticalised into a future marker affix -ta-. We see then that a future tense phrase consists of the following sequence:

```
```

Compare Dahl’s example from Swahili wa-ta-ku-dja (3p-FUT-INF-come) ‘they will come’.

Dahl (1954:358) argued that, when Malagasy adopted the Bantu tense system, it did not need the combination of a future tense auxiliary *caka and an infinitive prefix *ku. However, in rationalising this future tense marking, it dropped *caka instead of *ku, and it reinterpreted the latter as the future tense marker. *ku became present-day Malagasy hu and was further reduced to h- in verbal prefixes. The regular change from *k to h indicates

---

This example is analytically straightforward. However, most other instances are more complicated. For instance, Swahili verbal derivations expressing tense, ku- only surfaces before monosyllabic roots: elsewhere it has disappeared; in Shinzwani, *caka and *ku have merged to tso; and so on (cf. Dahl 1954:358).
that *ku must have been borrowed at a time not too long after the arrival of South East Barito speakers on the western shores of the Indian Ocean (cf. §2.1).

Finally, as already hinted on above, Dahl (1951) traced the Malagasy past tense marker back to the perfect aspect marker in Proto Austronesian. However, he also pointed out that while n- was inherited, the proclitic marker nu- came into being at a later stage and must have developed by analogy with hu-.

2.3 The agent prefixes mpaN- and mpi-

Malagasy has two prefixes mpaN- and mpi- which express the agent of an action, e.g. man-anatrà ‘to advise’ and mpananatrà ‘advisor’; mi-anatrà ‘to study’ and mpianatrà ‘student, pupil’. These prefixes are productive and occur with Austronesian as well as Bantu roots, cf. ma-musavi ‘to bewitch’ and mpa-musavi ‘witch’, cf. Swahili mchawi ‘witch’, uchawi ‘witchcraft’; mi-fidi ‘to choose’ and mpi-fidi ‘voter’ < PAn *piliq ‘to choose’; manuratrà ‘to write’ and mpanuratrà ‘writer’ (root: suratrà) < Malay surat ‘something written, letter, document’. The initial cluster mp- would be unusual in Malagasy phonotactic structure, but it is pronounced as [p]: the m- is not pronounced and the prefixes are realised as [paN]- and [pi]-. This initial mp- sequence is not an orthographic convention but is historically motivated. Regional Malagasy dialects such as Sakalava and Tanala have the corresponding disyllabic prefixes ampaN- and ampi- (Dahl 1954; Beaujard 1998). Ancient documents from the south and south-east of Madagascar have umpaN-, a variant form also attested in archaic forms of Sakalava (cf. umpan-dzaka ‘king’). The reduction of homorganic nasal + voiceless stop clusters to their voiceless stop component is a widespread trend in east Malagasy dialects, not only on morpheme-boundaries but also within roots.

MpaN- and mpi- do not reflect the Proto Austronesian prefixes *paN- and *paR-. These latter did survive in Malagasy but their regular reflexes are respectively faN- and fi-, noun-forming prefixes that do not normally refer to agents. Their meanings include ‘the way an action is usually performed’, ‘being in the state expressed by the verb’ (man-deha ‘to go’, fan-deha ‘manner of walking’), objects (mi-sutru ‘to drink’, fi-sutru ‘a drink’), instruments (mi-hugu ‘to comb’, fi-hugu ‘comb’), and (seldom) ‘one who performs an action as a habit’ (mi-tumani ‘to cry’, fi-tumani ‘cry-baby’).

Bantu languages make a clear morphological distinction between nominal derivations referring to agents (using the Bantu prefix *mu-) and other nominal derivations. According to Dahl, Malagasy may have developed mpaN-/mpi- as a new series of agent prefixes in order to compensate for the lack of Bantu *mu-. He speculated that mpaN- and mpi- may originally be concatenations of *mu- and faN-/fi-.

However, there are more straightforward solutions to the mystery of their origin. One possibility is that they historically consist of a shortened form of ulunà ‘person’ followed by faN- or fi-. This would explain why mpaN- and mpi- occurs dialectally as ampaN-/ampi- and has the archaic variant umpaN-. The reduction of ulunà is also seen in the variant forms n/devu, an/devu and un/devu ‘child; domestic, slave’ which historically

---

10 Incidentally, Swahili also has an aspect marker hu- denoting habitual action (Schadeberg 1992:38), but, while its form is identical to the Malagasy future marker hu, its meaning clearly shows that they are not related.

11 Prof. Noel Gueunier (email communication).
derived from *ulun + *lewu ‘house’ (Dahl 1951:312), and in the ‘Arabico-Malagasy’ term un/isinava ‘slave, prisoner’ deriving from *ulun *t<in>awan ‘captured person’, cf. Maanyan nawan ’to go to war, take prisoner’, and wuah tawan ‘captured, imprisoned’ (Dahl 1951:333). Other instances (provided to me by Professor Gueunier) are um/be ‘meritorious person, nobleman’ (< ulunǎ + be ‘big’); um/pe ‘blind person’ (< ulunǎ + pe ‘blind’), umarari ‘sick person’ (< ulunǎ + marari ‘ill’); possibly also ethnonyms such as an/ta/ndrui ‘Tandroy people’ which also occurs with initial u in the set phrase manalake amin-un/ndrui (take from-people/Tandroy) ‘take [a wife] from the Tandroy people’. In all these cases, the an-, -un- or initial nasal can be explained as a reduced form of ulun.

The vowel variation in an- and un- is due to the Malagasy tendency to neutralise historical antepenultimate high vowels to a. Note also that Malagasy non-stressed vowels tend to be reduced, which would affect the two last vowels in ulun.

It may seem odd to analyse mpaN- and mpi- as combinations of ulunǎ + faN-mpi-, judging from the meanings faN- and fi- have today, which do not usually refer to agents. However, historically the semantic roles of these forms must have been more open than they are today and may have included that of agent. In fact, it must have been the development of mpaN-mpi- as a new series alongside faN-mpi- that triggered the present-day role distribution between the two pairs of prefixes. Proto Austronesian *paN- and *paR- originally derived verbs. These verbs could also occur as nouns, as the capacity to function as a noun was a property of all Proto Austronesian verbs. The semantic role of nouns derived from *paN- and *paR- must have been very open, and it must have been largely dependent on context and on the semantics of the host root. In various present-day Austronesian languages (including Philippine and Batak languages), *paN- and *paR- still derive verbs, although there is also a tendency for certain derivations to become lexicalised as nouns. In other Austronesian languages, like Malagasy, Malay and modern Javanese, they have typically become nominal prefixes.

The development of Malagasy agent prefixes is to some extent similar to the development of Malay the agent-instrument prefixes pəN- and pəR-. The latter also derived from Proto Austronesian *paN- and *paR-. They went through a Proto Malayic stage, in which they formed deverbal nouns, the semantic role of which was basically determined by the context in which they occurred. These deverbal nouns would often be used attributively in noun phrases requiring a head which designated a person or a instrument. Over time, they ended up occurring as noun phrases by themselves (that is, without a head) and became agentive nouns or instrumental nouns (Adelaar 1992:179–193). Compare the following derivations:

- oraŋ ‘person’, (mə)mabantu ‘to help’ → oraŋ pəm-bantu ‘person who helps’
- → pəm-bantu ‘assistant’
- tali ‘rope, tape’, (məŋ-)ukur ‘to measure’ → tali pəŋ-ukur ‘measuring tape’
- → pəŋ-ukur ‘meter, gauge’; (also ‘someone who measures’).

It appears that Malagasy and Malay followed similar pathways in creating nominal affixes from the original Austronesian prefixes *paN- and *paR-. Although the developments in both languages are by no means identical, there is a clear parallel between, on the one hand, traditional Malay constructions such as oraŋ pəm-bantu yielding modern Malay pəmbantu through ellipsis of the nominal head, and on the other hand,
Malagasy constructions such as *ulunā *fananatrā which must have been the origin of modern Malagasy mpananatrā through phonological reduction of the nominal head, as postulated above. The likelihood that the ‘silent’ m in mpaN- and mpi- derives from an earlier *ulunā is increased by the fact that these prefixes are typically agent prefixes, and their meaning is unequivocal. By comparison, Malay paN- and pař- have a wider meaning, also deriving instrumental nouns. Moreover, some Malay paN- and pař- derivations are neither agent nor instrument and appear to be anomalies reflecting an earlier stage when the roles of paN- and pař- forms had not yet crystallised into their present-day ones; compare sakit ‘ill’ and pañakit ‘disease’ and dapat ‘to be able’ and pañdapat ‘opinion’.

This possibility has a greater explanatory force than Dahl’s theory that mpaN-/mpi-developed through Bantu influence. The latter many only account for the assumed need of agent markers as a semantic category. As far as their shape is concerned, however, these markers do not correspond to any agent prefixes in Bantu languages. If Bantu influence played any role at all, it was at best a concomitant one adding to the motivation to form prefixes that refer exclusively to agents.

Another, less likely, possibility is that mpaN- is a reflex of the Malay agentive-instrumental prefix paN-. The vowels agree: in Malagasy, antepenultimate schwa became a, and since most verbal roots have two syllables with a possible additional whispered vowel (CVCV(Cā)), mpaN- as a prefix is as a rule an antepenultimate syllable. However, whereas mpaN- conceivably reflects Malay paN-, it is unlikely that mpi- would also be a Malay loan morpheme, because there is no Malay form that matches it phonologically. On the other hand, assuming that mpaN- was indeed borrowed from Malay, it would be conceivable that mpi- were created in analogy to both Malagasy mpaN- and fi-. In this way, it would fill a morphological gap and satisfy the urge to restore paradigmatic balance in the affix system, as follows:

<table>
<thead>
<tr>
<th>Actor-oriented prefix</th>
<th>General nominalising prefix</th>
<th>Agent prefix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitive verbs</td>
<td>maN-</td>
<td>mpaN-</td>
</tr>
<tr>
<td>Verbs low in transitivity</td>
<td>mi-</td>
<td>fi-</td>
</tr>
<tr>
<td>(mpi- →) Ø</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A weakness of this scenario is that it does not explain why the agent prefixes have an mp-cluster. Although modern Malagasy does have a p, historical *p became Malagasy f, and words borrowed from Malay reflect *p as f (except in clusters with preceding m). The mp-cluster could therefore be explained as a phonological adaptation made at a time when *p had already become f and Malagasy had not yet its present-day p. However, this explanation remains ad hoc and is less attractive than the one involving a grammaticalised form of *ulun.

3 Malay influence

3.1 The passive prefix vua-

Merina has a passive prefix vua-, which marks actions that are accomplished and carried out by a conscious agent. Rasoloson and Rubino (2005:476) prefer to call vua- a resultative prefix, which encodes completive states with experiencers (non-agents) in

---

14 The only exception I found is puki ‘swearword, admonition’ (Gueunier 1986). It occurs in Comoran Malagasy and may be related to Malay puki ‘vagina’, which is used in insults.
subject function. The agent is usually not expressed; if it is, it immediately follows the verb prefixed with *vua*- (Rasoloson and Rubino 2005:477). Examples:

\[
\begin{align*}
Vua-vuri & \ni \ hetra. \\
& \text{PASS-collect ART tax} \\
& \text{‘The tax has been collected.’/‘The taxes have been collected.’ (Malzac 1960:58)} \\
Efa & \ vua-laza-ku \ t-\ami-nau \ sahadi \ izani. \\
& \text{PRF PASS-say-1s.POSS PST-with-2s.POSS already that} \\
& \text{‘I already told you that.’ (Malzac 1960:58)}
\end{align*}
\]

There is also a free verbal form *vua* meaning ‘hit, affected’. For example:

\[
\begin{align*}
Vua & \ ahu, \ \fa \ \mati \ ni \ \rai-ku. \\
& \text{hit 1s because dead ART father-1s.POSS} \\
& \text{‘I’m very sad [‘struck by misfortune’] because my father died.’}
\end{align*}
\]

Cognate to Merina *vua*-/*vua*, Sakalava has a verb *vua* ‘correct, hit the mark’. For example:

\[
\begin{align*}
Vua & \ zai \ volaŋ-enau \ zai. \\
& \text{correct REL word-2s.POSS REL} \\
& \text{‘What you say is correct.’}
\end{align*}
\]

Maanyan has a corresponding *wuah* with the same meaning. For example:

\[
\begin{align*}
Wuah & \ leŋq-nu \ jero. \\
& \text{correct word-2s.POSS that} \\
& \text{‘What you say is correct.’}
\end{align*}
\]

Maanyan *wuah* also occurs proclitically before a noun where it has the meaning ‘hit, affected’, e.g. Maanyan *wuah uran* ‘caught by the rain’; *wuah antah* ‘struck by fate’; *wuah tawanaw* ‘made a prisoner of war’.

Dahl (1951:214, 334) treated Merina *vua/vua*-, Sakalava *vua* and Maanyan *wuah* as regular reflexes of Proto Austronesian *buaq* ‘fruit’. He argued that Merina *vua-* was the latest stage in the grammaticalisation process from *buaq* ‘fruit’ to a passive prefix, whereas Sakalava *vua* and Maanyan *wuah* represented an intermediary stage.

However, tracing *vua*/vual/wuah* to Proto Austronesian *buaq* is problematic. The final *h* in Maanyan *wuah* is irregular, as Proto Austronesian final *q* normally became *ʔ* or Ø in South East Barito languages, and not *h*. It suggests borrowing from Malay, because final *q* as a rule became Malay *h*, compare Malay *buah* ‘fruit’ (also used metaphorically in the sense of ‘result’). The phonological irregularity in *vua*/vual/wuah* is underscored by the fact that Malagasy and Maanyan also have doublets that do reflect the original Proto Austronesian etymon *buaq* in a regular way from both a phonological and semantic point of view, i.e. Malagasy *vua* ‘fruit’, and Maanyan *wu?* ‘fruit’. Ngaju, which is spoken to the west of the Maanyan-speaking area,\(^\text{16}\) shows a parallel development as in Maanyan: it has *bua?* ‘fruit’ alongside *buah* ‘good, well, correct, fitting’, which is also used proclitically before verbal or nominal roots in constructions meaning ‘hit, reached by [root]’. Here

\(^{15}\) It is not clear whether or not *ʔ* is a reflex of final *q*. Its widespread occurrence in south and west Borneo (including after historical final vowels) also allows it to be interpreted as a (relatively recent) areal feature.

\(^{16}\) Ngaju (also Ngaju Dayak) is a West Barito language and an important contact language in the Barito area. However, contrary to what is suggested by their name, West Barito languages do not seem to be directly related to East Barito languages (Hudson 1994:16).
again, bua? ‘fruit’ is phonologically regular, whereas the -h in buah betrays borrowing from Malay. In Malagasy, the distinction between inherited and borrowed reflexes of *buaq may also have existed but cannot presently be made because all historical glottals were lost.

So, if Dahl was right in tracing vua/-vua/vuah to Proto Austronesian *buaq, this can only have happened indirectly via borrowing from Malay on account of the irregular final h in Maanyan vuah (and Ngaju buah). Both Maanyan and Ngaju are neighbouring languages and have been in contact with Banjar Malay for a very long time. The latter has been a major linguistic influence on the West and East Barito languages. The presence of Banjar Malay loanwords in Malagasy shows that Banjar Malay influence on South East Barito languages predates the Malagasy migrations to east Africa (Adelaar 1989, 2009). It is therefore likely that Malagasy vuah/vuah-, Maanyan vuah and Ngaju buah began their evolution as a Banjar Malay loanword, which spread through the West and East Barito area (see Map 2) while altering its original meaning from ‘fruit’ and ‘result’ to ‘correct, hit the mark’ and becoming an adversative adverb.

3.2 The prefix tafa-

The Merina prefix tafa- expresses an accomplished act. It loses its final -a before roots beginning with a or i. Example:

\[
\text{Kanefa mi-ezaká izahai ka tafiditrá \{tafa-iditrá\}.}
\]

‘But we hurried and managed to get in.’

Tafo- constructions can be non-volitional (e.g. mi-hauná ‘to meet’ vs tafo-hauná ‘to meet by accident’), but this is by no means always so, as can be seen in the previous sample sentence, where the agent is basically in control of the action. Tafo-constructions are usually actor-oriented, although undergoer-oriented cases also occur. The following sample sentence is patient-oriented and expresses non-volition. For example:

\[
\text{Tafandri \{tafa-andri\} t-ani ambanivuhitrá izi.}
\]

‘He spent the night in the country.’ (Malzac 1960:59)

Tafa- has been labelled in different ways. Many sources (Malzac 1960:56; Montagné 1931:50–55; Rajemisa-Raolison 1971:96; Rajaonarimanana 2001:52) consider it an undergoer-oriented prefix because of its morphological valency and (apparently) the undergoer orientation of some of the sentences in which it occurs, but this is obviously not accurate. Rasoloson and Rubino (2005:476–477) call it a resultative prefix because the act in question is implied to be accomplished and successful. In many ways tafa- seems to
express potentine aspect. However, although these latter descriptions are clearly more accurate, they also fail to capture the exact meaning of tafa-

Maanyan evidence for tapa-tapi-tapo- is minimal. Dahl’s one example, which is moreover out of context, is insufficient, and other grammatical sources for Maanyan do not mention these prefixes. On the other hand, derivations with non-volitional tapa- are also found in Hardeland’s Ngaju dictionary (1859), suggesting that this prefix has a larger spread in southern Borneo. Also, as in the case of vua- and its corresponding forms, tafa- and its cognates most likely have a (Banjar) Malay origin. Banjar Malay tapa- is an unproductive prefix expressing non-volitionality and occurring with verbs that otherwise take the intransitive prefix ba-, e.g. ba-hurup ‘to buy; to exchange’ and tapa-hurup ‘exchanged by accident’; ba-hual ‘to be quarrelling’ and tapa-hual ‘to get into a fight (unasked for)’ (Hapip 1977). It is a combination of ta- and pa-. The morpheme ta- is the default marker for non-volition in transitive as well as intransitive verbs. It is still productive and is related to Standard Malay te(r)-, which is also non-volitional (< Proto Malayic *tAr- ‘non-volitional prefix’). The morpheme pa- forms transitive verbal stems and is also found in deverbal nouns; its intransitive counterpart is ba-. It is related to Standard Malay pe(r)- and derives from Proto Malayic *pAr-, both of which have the same functions (Adelaar 1992:Ch.6).

The arguments for a Banjar Malay origin of tafa- and its cognates can be summarised as follows. There is no (Proto Austronesian) historical precedent for tafa-tapi-tapo-. Furthermore, Banjar Malay has long been a strong influence in south Borneo, long enough to exercise lexical influence on the language of the early Malagasy migrants to east Africa (cf. Adelaar 1989, 2009). Finally, tapa- can be explained as a combination of two other prefixes within the Banjar Malay morphological system.

4 The honorific personal prefix ra-

Malagasy has a fossilised prefix ra- which occurs in nouns with human reference, including kinship terms and proper names, e.g. rafutsi ‘term of address for old lady’ (< ra-+ futsi ‘white’); Comoro Malagasy ravinantu ‘child-in-law’ (< ra-+ vinantu < PAn *b-in-antu ‘child-in-law’); Ranavalona (name of a series of 19th century Malagasy queens), Rantoandro, Rajaonarimanana, Razafintsalama (Malagasy last names).

This prefix was adopted relatively recently in Malagasy, considering that other South East Barito languages do not have it. This is also apparent from internal evidence within Malagasy itself. Malagasy historically has two personal prefixes: apart from ra-, there is also z (sometimes iz-) deriving from early (pre-)Malagasy *i-,*y-, which in turn derives from the South East Barito person marker *i. It is attested as i- in South Betsimisaraka, Tanala and some other Malagasy dialects (Dahl 1951:56; Beaujard 1998). It is demonstrated in the following examples:

Zanakā ‘child, offspring’ < *y-anak < *i anak
Izahu, Tanala iahu ‘1s (emphatic)’ < *i aku
Izahai, Tanala iahai ‘we (excl.)’ < *i akai
Zandri, Tanala iandri ‘younger sibling’ < *i andri
Zafi ‘grandchild’ < *i ape
The prefix *z/*iz- reflects the Proto Austronesian personal article *i. In some exceptional cases it co-occurs with ra-: it then takes the form of -z- only and is closer to the root than ra-, as in razandri (< *ra- + *i + *andri) ‘younger sibling’.

Consequently, ra- must be borrowed, and more particularly from Javanese, which had a great impact on many languages in western Indonesia, including Banjar Malay. It also had an influence on Malagasy, probably at a time before its speakers migrated to east Africa (Adelaar 1995b, 2009). Zoetmulder (1982) gives Old Javanese ra-, a honorific prefix which occurs in titles and kinship terms, for example, rahadyan ‘Lord, Master’ (modern Javanese raden, < ra- + hadi + -an < Proto West Malayo Polynesian *qaji ‘ruler’). The Javanese prefix ra- has a full form sira-, with which it alternates in some forms. Furthermore, if (si)ra- is prefixed to a root beginning with a vowel, its vowel and the following vowel are contracted to a long vowel, for example, rāma, sirāma ‘father’, modern Javanese rāmå ‘Roman Catholic priest’ (< Proto Austronesian *ama ‘father’), and rānak, sirānak ‘son, daughter’ (< Proto Austronesian *anak ‘offspring’). This sira- is a cliticised form of the third person pronoun sira, sometimes also functioning as a second person pronoun (Zoetmulder 1982).

5 The ethnonymic prefix ta-

Another nominal prefix is ta-, which is sometimes directly followed by a nasal ligature -N- and refers to an ethnic or geographical group, or to someone who belongs to an ethnic group, resides in a certain place, or belongs to a certain profession. Compare the following examples: ta-lautrā ‘Talaotra, an ethnic group supposedly originating from across the Mozambican channel’ (cf. Malay laut ‘sea’); ta-nusi ‘Tanosy, an ethnic group in South Madagascar’ (lit. ‘island people’, cf. Malagasy nusi ‘island’); ta-n-tsaha ‘farmer, someone living in the fields’ (saha ‘rice field’); ta-m-buhitrā ‘those who live in the hills’ (vuhitrā ‘hill’); ta-n-dapa ‘people who live in, or frequent, the palace’ (lapa ‘court; main residence; pavilion in the centre of a village’); ta-m-bi ‘blacksmith’ (vi ‘iron’); ta-m-batu ‘stoneworker, mason’ (vatu ‘stone’).17

The prefix ta- ultimately derives from the Proto Austronesian word *taw ‘human being, person’. However, in Malagasy as well as in other East Barito languages, *taw was replaced by *hulun as the default form for ‘human being’. It has left no trace, except for the ethnonymic prefix in Malagasy (Adelaar 1995a). Malagasy ta- could be a relic form directly inherited from Proto Austronesian, but its use as a prefix and its specific meaning also suggest borrowing as an alternative explanation. In South Sulawesi languages (as well as in languages in other parts of Sulawesi), this etymon survived as a free form *taw (the default term for ‘human being, person’) and as a prefix ta-, to- or tu- (cf. To-raja lit. ‘people from the inland’, the name of an ethnic group). As South Sulawesi languages did exercise influence on Malagasy, and the prefix is not found in any of the other Austronesian lending languages (Malay, Javanese, Ngaju), the chances that ta- was borrowed from a South Sulawesi language are real (Adelaar 1995a).

6 Possible origins of the undergoer-oriented imperative suffix -u

The emergence of new affixes is of course not always triggered by language contact. For instance, the Malagasy imperative ending -u may be the result of grammaticalisation of

17 With thanks to Prof. Noel Gueunier for the last five examples.
an inherited modal particle. It is suffixed to non-agent-oriented verbs\(^ {18}\) that do not have an \(u\) in the last syllable of their root (e.g. the root \(hafa\)trā ‘message; recommendation’ forms a verb \(mannafa\)trā ‘to entrust someone with a commission, order merchandise’, which has an undergoer-oriented form \(hafa-an\)ā and an imperative undergoer-oriented form \(hafa-u\) ‘That it be sent in commission! That it be ordered!’). It has an allomorph \(-i\), which is suffixed to verbal roots that have \(u\) in their last syllable (e.g. \(sulu\), \(manulu\) ‘to replace’ has an undergoer-oriented form \(sulu-an\)ā, and an imperative undergoer-oriented form \(sulu-i\) ‘Replace!’). Bayan, one of the South East Barito languages,\(^ {19}\) has a modal article \(-hu\) ‘just, only’. When it is cliticised to a verb, it adds emphasis to it, and in some cases the result is an imperative phrase. Compare in the following examples: \(tulak-hu Mahaji\) (leave-MODAL Mahaji) ‘(So) Mahaji left!’ (Nanang et al. 1988:81); \(okan-hu dadeo!\) (eat-MODAL much) ‘Eat well!’; \(ugo-hu ika utek-ne\) (take-MODAL 2s head-3s.POSS) ‘(just) take its head!’ (Nanang et al. 1988:102). It would be worthwhile to obtain more information on Bayan \(-hu\) and corresponding forms in other South East Barito languages in order to investigate if they are related to the Malagasy passive-imperative suffix \(-u\). For the time being, however, this remains speculative in view of the minimal information available on Bayan \(-hu\) and the apparent lack of cognates in other South East Barito languages. Moreover, this is not the only possible explanation for the imperative allomorphs \(-u\) and \(-i\). Ross (1995:742–743), reconstructs for Proto Austronesian a series of atemporal suffixes (actor-oriented *-Ø, undergoer-oriented *-u, and locative-oriented *-i). In the Austronesian daughter languages that have reflexes of these atemporal suffixes, these reflexes have (one or several of) three basic functions: (1) they express the imperative; (2) they express events in sequence in narrative; (3) they derive forms that occur subordinate to some auxiliaries. However, the reconstruction of this series is not without its problems. There are only three Formosan languages (Saara, Puyuma, Paiwan) that exhibit a reflex of *-u, and (according to Ross), there is only one extra-Formosan (i.e. Malayo-Polynesian) reflex of *-u, namely in Lun Daye, a language of the Apo Duat subgroup in Sarawak, north Borneo. In the latter, \(-u\) and \(-a\) are both imperative undergoer-oriented suffixes, \(-u\) being used for actions close to the speaker, and \(-a\) for actions at a distance from the speaker (Ross 1995:763). Ross does not mention the Malagasy the imperative suffixes \(-u\) and \(-i\). As discussed above, these are attached to any verbs that are not actor-oriented, and the difference between them is phonologically determined. Furthermore, the distribution of Malagasy \(-u\) and \(-i\) is rather different from the distribution of reflexes of *-u and *-i elsewhere. Malagasy \(-i\) probably simply reflects the Proto Austronesian subjunctive marker *-a, since Proto Austronesian final *a and *i regularly merged to Malagasy \(i\). Nevertheless, there is a remote possibility that they reflect *-u and *-i. At this point, neither Proto Austronesian *-u and *-i nor Bayan \(-hu\) provide explanations that are decisive for the origin of the Malagasy non-actor-oriented suffixes \(-u\) and \(-i\), but at least they are important leads for further research.

7 Concluding remarks

This chapter gave an overview of the way in which Malagasy integrated various new affixes into its morphology after it had split off from other South East Barito languages.

---

\(^ {18}\) Including patient-oriented verbs and the so-called ‘circumstantial’ verbs, which can take various other parts of speech as subject.

\(^ {19}\) Spoken in the vicinity of Muara Teweh, Barito Utara Regency, in central Kalimantan.
These affixes came from various sources. The nominal prefixes *ki- and *tsi- as well as the future tense markers *hu and *h- are of Bantu origin; the verbal prefix *tafa- must be borrowed from Banjar Malay, and *vuah-vua may have originated as a lexical borrowing from (Banjar) Malay, and possibly via Ngaju. The agent prefixes *mpaN- and *mpi- probably developed from a combination of *ulun + the nominal prefixes *faN- and *fi-; the now fossilised honorific person prefix *ra- must be borrowed from Javanese, and the ethnomic prefix *ta- is probably of South Sulawesi provenance, although it could also be a retention of Proto Austronesian *taw 'person'. Finally, two possibilities are proposed for the origin of the undergoer-oriented imperative suffix -u: it is either a reflex of a Bayan (South East Barito) modal particle *hu-, or, together with -i, it reflects the Proto Austronesian atemporal undergoer-oriented suffixes *-u and *-i.

The prefixes originating from Bantu were obviously adopted after South East Barito speakers had migrated to east Africa. The prefixes *mpaN- and *mpi- are probably also post-migratory, as they seem to be fairly recent developments. As to the prefixes *vuah- and *tafa-, their ancestral forms were already borrowed before the migrations, as they also occur in other Barito languages and their source language must be Banjar Malay. This was also probably the case with *ta-, if this is indeed a South Sulawesi loanword, because contacts between South Sulawesi languages and Malagasy in all likelihood happened in the South East Barito homeland, which in navigational terms is relatively close to South Sulawesi (Adelaar 1995a, 2009). Whether the imperative suffix -u can be traced to South East Barito or to Proto Austronesian, it has been in Malagasy since pre-migratory times. The originally Javanese prefix *ra- is clearly post-Proto South East Barito, as it does not occur in other Barito languages. It was most likely borrowed into early Malagasy at some stage immediately preceding the migrations to East Africa.\footnote{Although theoretically this borrowing might also have occurred directly after the migration, depending on whether the Javanese were also involved in transoceanic contacts, and to what extent. However, there is as yet no clear evidence to suggest that they were.}

It appears that Bantu morphological influence was probably not as fundamental as implied in Dahl’s work. Some of the prefixes that Dahl assumed to be very old seem to be of a relatively recent date, and some prefixes that he attributed to a Bantu substratum are more readily explained as due to an internal development or to Malay influence.

The affixes discussed here may not be the only later additions to the original morphology of Malagasy. Further research is required to obtain a more comprehensive insight into the grammatical history of Malagasy.

References


The amalgamation of Malagasy


1 Introduction

Proto Oceanic had a morpheme *akin[i] which introduced participants with particular semantic roles into a clause (see Pawley 1973; Harrison 1982; Evans 2003). *Akin[i] denoted different types of participants with different semantic classes of verbs. Table 1 shows the proposed correlations between the type of verb and the type of participant introduced by Proto Oceanic *akin[i].

Table 1: Types of participants denoted by Proto Oceanic *akin[i]

<table>
<thead>
<tr>
<th>Type of verb</th>
<th>Role marked by *akin[i]</th>
</tr>
</thead>
<tbody>
<tr>
<td>motion verbs</td>
<td>concomitant</td>
</tr>
<tr>
<td>psychological and emotional states</td>
<td>cause/stimulus</td>
</tr>
<tr>
<td>speech and cognition</td>
<td>content</td>
</tr>
<tr>
<td>excretion/secretion</td>
<td>product</td>
</tr>
<tr>
<td>process-action verbs</td>
<td>instrument, benefactive</td>
</tr>
</tbody>
</table>

Examples (1) to (6) demonstrate each of these uses of Proto Oceanic *akin[i] with its reflexes in contemporary Oceanic languages. In (1), from Longgu (SES), the motion verb *ango 'crawl' takes the suffix -ta'ini to introduce a concomitant participant as the direct object. In (2) Woleaian (Mic) *yagili occurs with an emotional state verb and introduces an...
oblique argument denoting the cause of the state. The North-East Ambae (NCV) suffix 
-gi(ni) with a verb of thought, as in (3), allows the content of the thought to be expressed 
as the direct object. Example (4), from Manam (NNG), demonstrates the use of the suffix 
-aʔ with an excretion verb with the product expressed as the direct object. Examples (5) 
and (6) show uses of *akin[i] reflexes with process-action verbs. In (5), from Tolo (SES), 
the preposition hini introduces an instrument participant, and in (6), from Ulithian (Mic), 
the preposition yixili introduces a beneficiary participant.

Longgu (SES)
(1) [Mwaa-i] A e ano-ta’i-i-r[_a] ni [gale ngai-gi].
snake-SG 3s crawl-TR-3p baby 3s-PL
‘The snake is crawling with its babies (on its back).’ (Hill 1992:59)

Woleaian (Mic)
(2) [I] S sa ker [yagili-g]PP.
1s PRF be.happy with-2s
‘I am proud of you.’ (lit: ‘I am happy with you.’) (Sohn 1975:287; gloss mine)

North-East Ambae (NCV)
(3) [Da] A =ni domi-gi [na problem ngihe]O.
1NSG.IN.SBJ=IRR think-TR ACC problem that
‘We should think about the problem.’ (Hyslop 2001:328)

Manam (NNG)
woman taro 3s.R-urinate-THC-TR-3p.OBJ
‘The (mythical) woman urinated taros (i.e. taros grew out of her urine).’ (Lichtenberk 1983:177)

Tolo (SES)
1s cut tree with-3s knife
‘I cut the tree with a knife.’ (Crowley 1986:14; gloss mine)

Ulithian (Mic)
kng DET 3s PRF need food for-3p soldier DET
‘The king needed food for the soldiers.’ (Sohn and Bender 1973:181)

The core arguments of each clause are enclosed in square brackets for ease of interpretation of valency-changing derivations. If a core argument is denoted by a noun phrase and a pronominal marker within the verb complex, then only the noun phrase will be labelled as the argument. However, if there is no noun phrase present in the clause then the pronominal marker will be labelled as the argument. This does not reflect any underlying assumptions regarding the argument status of noun phrases and the pronominal markers within the verb complex, but is simply a convention of presentation.
As these examples show, Proto Oceanic *akin[i] is reflected as both a preposition
(Woleaian, Tolo and Ulithian) and a transitivising affix (Longgu, North-East Ambae and
Manam). This overview of reflexes of *akin[i] is suggestive of a straightforward historical
scenario of a free form becoming an affix. However, such an analysis is problematic when
particular details of *akin[i] reflexes are taken into account. In this paper a more realistic
reconstruction of *akin[i] is attempted by considering the history of *akin[i] and its
reflexes from the perspective of change in progress.

In §2 four representative reflexes of *akin[i] in Oceanic languages are described from
the perspective of their degree of phonological independence and their grammatical status;
the two aspects of Proto Oceanic *akin[i] which are considered in §3 and §4, respectively.
In §5 it is concluded that Proto Oceanic *akin[i] was undergoing two processes of change.
First, in Proto Oceanic there was variation in the degree of phonological independence of
*akin[i]: with some verbs *akin[i] was phonologically dependent and with others it was not.
Since the break-up of Proto Oceanic, the number of verbs with which *akin[i] was
phonologically bound has gradually increased. Second, Proto Oceanic *akin[i] probably
had two structural analyses: an original one as a post-verbal modifier within the verb
complex, and an innovative one as the head of a prepositional phrase, both of which are
reflected in contemporary Oceanic languages.

2 Reflexes of *akin[i] in Oceanic languages

Proto Oceanic *akin[i] is widely reflected in contemporary Oceanic languages as a
transitivising suffix and a verbal preposition. For example, the reflex of *akin[i] in Manam
(NNG) is the transitivising suffix -aʔ (Lichtenberk 1983:222–240). Examples (7) and (8)
show intransitive and transitive uses of the verb nanari ‘to tell a story’. In (8) -aʔ has an
applicative function, allowing the content of the speech act to be expressed as the object
argument.4

(7) [U]-nanári.
1s.r-tell.story
‘I told a story.’ (Lichtenberk 1983:138)

(8) [Yábu]-nanarí-t-aʔ-i.
Yabu 1s.r-tell.story-THC-TR-3s.OBJ
‘I told the story about Yabu.’ (Lichtenberk 1983:138)

Manam -aʔ also has a valency-rearranging use, as shown in (9) and (10). The transitive
clause in (9) has a patient participant expressed as the object argument, however, in (10)
where the verb occurs with the transitivising suffix -aʔ, it is an instrument participant
which is expressed as the object argument.

(9) [ʔáti]-rózo-ŋ-i.
canoe 1s.r-plug.up-THC-3s.OBJ
‘I plugged up the canoe.’ (Lichtenberk 1983:238)

(10) [ʔáti]-rozó-ŋ-aʔ-i.
stick 1s.r-plug.up-THC-TR-3s.OBJ

4 Manam -aʔ also has a causative function, but this will not be discussed further here.
‘I plugged the stick in.’ (Lichtenberk 1983:238)

As demonstrated by examples (8) and (10), participants introduced by -aʔ are expressed as core, and not oblique, arguments. In (8) the ‘content’ participant, Yabu, occurs without any following case suffix or post-position and is indexed by the 3s object suffix -i, both characteristics of object arguments (Lichtenberk 1983:158, 170–172).

The phonological dependence of -aʔ is apparent from the placement of stress, such that primary stress is assigned on the basis that a sequence of a verb plus -aʔ forms a single phonological word (Lichtenberk 1983:50–84). That -aʔ forms a phonological word with the preceding verb stem is also seen from the presence of thematic consonants between the verb root and -aʔ which only occur before non-zero suffixes (Lichtenberk 1983:136–158).

In other languages, like Tamambo (NCV), reflexes of *akin[i] are somewhat different in terms of both their grammatical status and degree of phonological independence. The reflex of *akin[i] in Tamambo is a preposition that is realised as hini or hina depending on the nature of its object. Hini ~ hina occurs in transitive and intransitive clauses, as in (11) and (12) respectively, introducing participants with a range of semantic roles expressed as oblique arguments (Jauncey 1997:61–62). Unlike in Manam, in Tamambo the reflex of *akin[i] is phonologically independent of the verb stem.

mum 3s give-1s.OBJ PREP some-LK tea
‘Mum gave me some tea.’ (Jauncey 1997:61)

3s-TA happy PREP PL-child-3s.POSS
‘He is proud of his children.’ (Jauncey 1997:62)

The reflex of *akin[i] in Boumaa Fijian, the transitive suffix -Ca'ina ~ -Ca'ini, has the grammatical status of an affix, but is phonologically independent.5 Dixon (1988:200–204) describes -Ca'ina as a transitive suffix with both applicative and causative uses. For example, with the verb dredre ‘to laugh’ -Ca'ina has an applicative function, deriving dredre-va'ina ‘to laugh at someone/something’, whereas with volo ‘to be hidden’ -Ca'ina has a causative function, deriving volo-ta'ina ‘to hide something’.

Affixes in Boumaa Fijian are morphemes which form a single grammatical word with their host and are defined as ‘a form which, when added to a root, changes the syntactic potential of that root’ (Dixon 1988:27). Since -Ca'ina affects the argument structure of the verb, Dixon (1988) analyses it as an affix.

However, -Ca'ina is not phonologically dependent on the preceding verb stem. This can be seen from the placement of stress. Primary stress in Boumaa Fijian occurs on the penultimate mora of each phonological word (Dixon 1988:17), and a verb stem and -Ca'ina both take primary stress, as in tōlanōa-ta'ina ‘to relate (a story)’ where primary stress occurs on the penultimate mora of the verb stem and the penultimate mora of the transitivising affix. In this way a verb and -Ca'ina can be seen to represent two phonological words.

Both the grammatical status and the degree of phonological independence of the Woleaian reflex of *akin[i], yagili, are ambiguous. Sohn (1975:284–287) describes yagili

5 In Boumaa Fijian the -Ca'ini form of the transitive suffix occurs if the object argument is expressed by a pronoun or a proper noun, while the -Ca'ina form occurs elsewhere.
as a verbal preposition, one of a small closed class of words which have features of both verbs and prepositions. Yagili is prepositional in nature in that it introduces an adjunct (rather than a core argument), but is verbal in nature in that it takes verbal object suffixes indicating the person and number of its object. Although Woleaian yagili is a preposition and is not considered to be part of the verb complex, it is closely associated with the verb. For example, certain verbs require that an adjunct be introduced with yagili, and not one of the other verbal prepositions. Yagili also tends to occur adjacent to the verb preceding any object noun phrase and verbal modifiers such as directionals, as in (13), and certain adverbs as in (14). Note that in (14) the object of yagili is expressed by the object suffix on the following adverb fengan ‘together’.

(13)  *Rig yagilii-r lag sar kela.*
run with-3p.OBJ away child DEM.PL
‘Run away with those children.’ (Sohn 1975:287)

(14)  Ye sa teo yagili fenganii-r tag.
3s PRF climb with together-3p.OBJ away
‘He has carried all of them up.’ (Sohn 1975:285)

The degree of phonological independence of yagili is variable. In Woleaian root-final vowels are voiceless in word-final position (and not written in the orthography), but are voiced when not in word-final position, as when followed by a suffix. In (15) ker ‘be happy’ occurs with the suffix -l and the root-final vowel e is voiced (and shows up in the orthography), whereas in (2) above ker ‘be happy’ occurs as [keʐe] with a voiceless form of the final vowel when followed by yagili, evidence that they form two separate phonological words.

(15)  Ye kere-l gan.
3s be.happy-of food:POSS
‘He is happy because of his food.’ (Sohn and Tawerilmang 1976:76)

However, with verb roots ending in -a, yagili occurs without the initial glide and forms a single phonological word with the verb root. Thus in (16) mmwut ‘to vomit’, which occurs as [mːʷureadystatechange] when not followed by a suffix, occurs as [mːʷutaɣiri], where the final vowel of the verb root is voiced and forms a long vowel with the initial vowel of agili.

(16)  Ye mmwuta-agili metta?
3s vomit-with what
‘What did he vomit?’ (Sohn 1975:287)

Thus contemporary reflexes of *akin[i] differ significantly from each other both in terms of their degree of phonological independence and their grammatical status, and it is these two issues that will be examined here with regards to Proto Oceanic *akin[i].

3 The degree of phonological independence of Proto Oceanic *akin[i]

The major issue concerning the degree of phonological independence of Proto Oceanic *akin[i] is the way it is reflected in the daughter languages with respect to the thematic consonants. Thematic consonants are lexically-determined consonants which occur between a lexical stem and a following suffix. Historically, thematic consonants reflect
original Proto Oceanic root-final consonants that have been lost in word-final position, but retained before a suffix. For example, the Proto Oceanic verb root *tanis ‘to cry’ is reflected in Kwaio as ani ‘to cry’, where the final *s of the Proto Oceanic verb has been lost. However, the transitive form of this verb, *tanis-i-, is reflected as ani-si- ‘to cry for someone’. In this form the original root-final consonant of Proto Oceanic has been retained because it was followed by the transitive suffix *-i, and it has been reanalysed as part of the suffix.\(^6\)

When considering *akin[i], it can be argued that if thematic consonants occur before its reflexes in a particular language then it must have been suffixed prior to the loss of word-final consonants. Thus the Kwaio verb ani-te’eni- ‘try to get something by crying about it’, reflecting Proto Oceanic *tanis ‘to cry’ and the Kwaio reflex of *akin[i], -Ce’eni-, with an historically correct thematic consonant, implies that *akin[i] was a suffix prior to the loss of word-final consonants. However, if thematic consonants never occur before reflexes of *akin[i] then it must have become a phonologically dependent after the loss of final consonants. For example, in Kiribatese (Mic) the verb stem tebo~tebo ‘to dive, to bathe’, has a transitive form tebo-ka ‘to pour water on something’, with a thematic consonant. However, the form with -akina, tebo-akina ‘to be insistent on something’ without the thematic consonant, suggests that in this language *akin[i] became phonologically bound subsequent to the loss of word-final consonants (Harrison 1982:185; Sabatier 1971).

Such conflicting evidence, where thematic consonants occur with some reflexes of *akin[i] but not others, led Harrison (1982:216–218) to suggest three possible historical scenarios for Proto Oceanic *akin[i]:

(a) that *akin[i] was originally a suffix and has been retained as such in languages like Kwaio which reflect *akin[i] with thematic consonants;

(b) that *akin[i] was originally phonologically independent and that the thematic consonants occurring before its reflexes in contemporary Oceanic languages do not reflect original word-final consonants; and

(c) that the loss of final consonants occurred independently in different groups of languages; before *akin[i] was phonologically dependent in some languages and after it was phonologically dependent in others.

It is indeed the case that the loss of word-final consonants occurred more than once in the history of Oceanic languages, with some languages or groups of languages showing a loss of word-final consonants and others a retention (see Ross 1988; Blevins 2004; Lynch 2005). However, a more detailed reconstruction of Harrison’s scenario (c) can be achieved by looking not simply at the presence or absence of thematic consonants with reflexes of *akin[i], but by also investigating how often and in which languages thematic consonants with reflexes of *akin[i] do indeed reflect original Proto Oceanic word-final consonants. This type of investigation is taken up first in §3.1 with Southeast Solomonic languages, where it is concluded that the reflex of *akin[i] in Proto Southeast Solomonic was likely to have been phonologically dependent with around half of the verbs with which it occurred. The survey of thematic consonants with reflexes of *akin[i] is broadened in §3.2 to

---

\(^6\) In the description of many Oceanic languages thematic consonants are analysed as forming part of the transitive suffix. Lichtenberk (1983, 2001) presents evidence against such an analysis for Manam and Toaqaqita (SES) and in support of an analysis of thematic consonants as a morph separate from both the transitive suffix and the verb stem. While it is likely this analysis applies equally well to other Oceanic languages, the analysis presented in the description of each language is retained here.
languages outside of the Southeast Solomonic subgroup, where it is argued on the basis of lexical reconstructions that *akin[i] was phonologically dependent with at least some verbs in Proto Oceanic.

### 3.1 Thematic consonants and *akin[i] reflexes in Southeast Solomonic languages

As mentioned above thematic consonants occur with the reflex of *akin[i] in Kwaio, and the same is true of all other Southeast Solomonic languages, suggesting that *akin[i] was phonologically bound in pre-Proto Southeast Solomonic before the loss of final consonants. More significantly, in Southeast Solomonic languages, reflexes of *akin[i] with some verbs can be shown to behave in a manner parallel to reflexes of *-i, such that the thematic consonants reflect original word-final consonants, as demonstrated by the data in Table 2.7

<table>
<thead>
<tr>
<th>Intransitive form</th>
<th>Form with *-i</th>
<th>Form with *akin[i]</th>
</tr>
</thead>
<tbody>
<tr>
<td>'cry'</td>
<td>*tanjis</td>
<td>*tanjis-i-</td>
</tr>
<tr>
<td>POc</td>
<td>*tanjis</td>
<td>*tanjis-i-</td>
</tr>
<tr>
<td>Gela</td>
<td>tanji</td>
<td>tanji-hi-</td>
</tr>
<tr>
<td>Kwaio</td>
<td>ani</td>
<td>ani-si-</td>
</tr>
<tr>
<td>Arosi</td>
<td>aghi</td>
<td>aghi-si-</td>
</tr>
<tr>
<td>'hit, kill'</td>
<td>*Ra(b,p)us</td>
<td>*Ra(b,p)us-i-</td>
</tr>
<tr>
<td>Longgu</td>
<td>–</td>
<td>rabu-si</td>
</tr>
<tr>
<td>Lau</td>
<td>rabu</td>
<td>rabu-si</td>
</tr>
<tr>
<td>Kwaio</td>
<td>labu</td>
<td>labu-si</td>
</tr>
<tr>
<td>'twist'</td>
<td>*pilos</td>
<td>*pilos-i-</td>
</tr>
<tr>
<td>PCEOc</td>
<td>filo</td>
<td>filo-si-</td>
</tr>
<tr>
<td>Lau</td>
<td>–</td>
<td>filo-si-</td>
</tr>
<tr>
<td>Kwaio</td>
<td>hiro</td>
<td>hiro-si-</td>
</tr>
<tr>
<td>Arosi</td>
<td>hiro</td>
<td>hiro-si-</td>
</tr>
<tr>
<td>'squeeze, press'</td>
<td>*peles</td>
<td>*peles-i-</td>
</tr>
<tr>
<td>PCEOc</td>
<td>fele</td>
<td>fele-si-</td>
</tr>
<tr>
<td>Lau</td>
<td>fele</td>
<td>fele-si-</td>
</tr>
<tr>
<td>Kwaio</td>
<td>fele</td>
<td>fele-si-</td>
</tr>
</tbody>
</table>

The correspondence between thematic consonants with reflexes of *akin[i] and reflexes of *-i also suggests that *akin[i] was predominately bound in Proto Southeast Solomonic. Thematic consonants which occur with -Ci in Southeast Solomonic languages tend to

---

7 In the Cristobal-Malaitan languages, a subgroup of Southeast Solomonic, Proto Oceanic *s is reflected as s before the high vowels i and u, and as t elsewhere. Thus in languages like Longgu, Kwaio, Lau and Arosi, what appear to be different thematic consonants before *-i and *akin[i] reflexes reflect a regular sound change, where *s has become t before the *akin[i] reflexes, but has remained as s before the *-i reflexes. The regular reflex of Proto Oceanic *p in S’a’a is h and the regular reflex of Proto Oceanic *s in Gela is h.
reflect the original Proto Oceanic word-final consonants. Thus, from a sample of 85 Proto Oceanic verbs reconstructed as consonant-final and occurring with *-i, 27 were reflected with -Ci in Gela (SES) and 20 (or 74%) of these occurred with the historically correct thematic consonant. For example, Proto Oceanic *lubat-i- ‘to loosen, untie’ is reflected in Gela as luba-ti- ‘to let something go, slacken something’ where the thematic consonant t reflects the Proto Oceanic word-final consonant *t. As can be seen from the figures in Table 3, in Kwaio it is 79% of the verbs reflected with -Ci which occur with the historically correct thematic consonant, whereas in Arosi it is only 51% of the verbs. This low percentage seems to reflect a tendency in Arosi for an innovative thematic consonant to occur where the original consonant would have been lost through a general sound change. Thus, of the 18 verbs with historically unexpected, or ‘incorrect’, thematic consonants in Arosi, 14 (or 78%) are lexemes which were *q-final or *t-final in Proto Oceanic, consonants that were lost universally in Arosi. For example, Proto Oceanic *kaRat-i- ‘to bite something’ is reflected in Arosi as ?ara-ʔi- ‘to bite something’ with an innovative ʔ thematic consonant. The two verbs in Arosi listed in Table 3 as showing an expected lack of a thematic consonant are verbs which were *t-final or *q-final in Proto Oceanic and are reflected in Arosi with the transitive suffix -i and no thematic consonant. Thus Proto Oceanic *asaq-i- ‘to grate, sharpen by grating or rubbing’ is reflected in Arosi as ata-i- ‘to scrape, sharpen by rubbing’.

Table 3: Thematic consonants with *-i in Southeast Solomonic languages

<table>
<thead>
<tr>
<th></th>
<th>Gela</th>
<th>Kwaio</th>
<th>Arosi</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of verbs reflected with *-i</td>
<td>27</td>
<td>19</td>
<td>41</td>
</tr>
<tr>
<td>no. of these verbs with ‘correct’ thematic consonant</td>
<td>20 (74%)</td>
<td>15 (79%)</td>
<td>21 (51%)</td>
</tr>
<tr>
<td>no. of verbs with expected lack of thematic consonant</td>
<td>0</td>
<td>0</td>
<td>2 (5%)</td>
</tr>
<tr>
<td>no. of verbs with ‘incorrect’ thematic consonant</td>
<td>7 (26%)</td>
<td>4 (21%)</td>
<td>18 (44%)</td>
</tr>
</tbody>
</table>

8 The Proto Oceanic reconstructions were taken from Ross, Pawley and Osmond (1998, 2003) and Evans (2003), and the Gela, Kwaio and Arosi data are from Fox (1955), Keesing (1975) and Fox (1978), respectively.

9 In Kwaio, where Proto Oceanic *t and *q are also universally lost, the reflex of *kaRat-i- is ?ale- ‘to bite something’. The tendency in Kwaio with originally *t-final verbs is for the final vowel of the verb stem and the *-i transitive suffix to merge, resulting in a class of verbs which have transitive forms that take the object suffixes directly, but which also have a different final vowel in the intransitive and transitive forms.

10 In some Oceanic languages default thematic consonants have emerged with reflexes of *-i and/or *akin[i]. Such generalisations of one or two thematic consonants within a language can lead to the occurrence of non-historical consonants with many verbs. Since thematic consonants with reflexes of *-i are rarely generalised in Southeast Solomonic languages, and tend to reflect original word-final consonants, innovative thematic consonants with -Ci are not taken to occur with more than a small proportion of verbs in these languages. The generalisation of thematic consonants with reflexes of *akin[i], such as in Longgu where only r and t occur as thematic consonants with -Ca’ini (Hill 1992:57), may have also led to innovative thematic consonants with reflexes of *akin[i]. In terms of the reconstruction of *akin[i] as phonologically dependent with particular verb stems, this would mean a more conservative reconstruction in the sense that some verbs which originally did occur with a reflex of *akin[i] and the historically correct thematic consonant may now have an innovative thematic consonant. It is assumed that this does not greatly alter the overall reconstruction of Proto Oceanic *akin[i].
Thus even in the case where there is no Proto Oceanic reconstruction, if thematic consonants with *akin[i] reflexes in Southeast Solomonic languages correspond with those with -Ci, it is likely that the thematic consonant reflects an historical word-final consonant that has been retained in both cases because it was not in word-final position at the time when the final consonant was lost.

Of the 119 verbs in the Gela dictionary which occur with -Caɣi, the reflex of *akin[i], 67 or 56% have the same thematic consonant with -Caɣi as with -Ci and are thus potential candidates for verbs with which *akin[i] was phonologically bound in Proto Southeast Solomonic. A few of these verbs have cognates in other Southeast Solomonic languages with corresponding thematic consonants, supporting the reconstruction of verbal forms with -Ci and -Caɣi for Proto Southeast Solomonic. These reconstructions and their supporting data are given in Table 4. While the reconstruction of these forms is well-supported, the reconstruction of the meanings is more difficult. For example, with Proto Southeast Solomonic *to[n,d]o-maɣi(ni), the Kwaio and Sa’a reflexes suggest a meaning little different from the transitive form with *-Ci, while the Gela reflex suggests a causative meaning.

| Table 4: Proto Southeast Solomonic verbs with *-Caɣi[n] |
|------------------|------------------|
| **Proto Southeast Solomonic** | **Gela** |
| *siki-li- * 'poke s.th. out' | **siki-laɣi(ni)-** |
| siki-li | 'poke out, pluck out, tickle' |
| **Arosi** | **Sa’a** |
| sigi-ri | 'poke out dirt' |
| sisiki-li- | 'flick stones with fingers' |
| **Proto Southeast Solomonic** | **Gela** |
| *susu-vi- | 'to suckle' |
| susu-vi | 'drink from breasts TR' |
| **Bughotu** | **Lau** |
| suuvi- | 'suck the breast' |
| **Kwaio** | **Lau** |
| susu-fi- | 'suckle' |
| **Proto Southeast Solomonic** | **Gela** |
| *to[n,d]o-mi- | 'swallow' |
| sono-mi | 'swallow' |
| **Lau** | **Kwaio** |
| ono-mi- | 'swallow' |
| **Sa’a** | **Sa’a** |
| ono-mi- | 'swallow' |

Data from Fox (1955, 1974, 1978); Ivens (1918, 1940); Keesing (1975)

For two of these verbs there is external (that is, non-Southeast Solomonic) evidence supporting the hypothesis that the thematic consonant reflects an original final consonant. Thus a Proto Oceanic reconstruction *to[n,d]om-i- 'to swallow' with a final *-m is well-
supported. While cognates in many Oceanic languages support the reconstruction of a form *susu ‘breast, to suckle’ without a final consonant, it is likely that the thematic *v in Proto Southeast Solomonic reflects the final consonant of an earlier form reconstructed for Proto Austronesian as *susup ‘suck’ (Blust 1995). With both these verbs it seems likely that *akin[i] was phonologically bound in Proto Southeast Solomonic before the loss of final consonants, explaining the presence of the historically correct thematic consonant with reflexes of *akin[i] in the contemporary Southeast Solomonic languages. The correspondence between thematic consonants with -Ci and -Caɣ with 56% of verbs in Gela suggests that *akin[i] was likely to have been phonologically bound with a similar proportion of verbs in Proto Southeast Solomonic prior to the changes that resulted in the loss of final consonants.

### 3.2 Thematic consonants with *akin[i] outside of Southeast Solomonic

The reconstruction of lexemes with *akin[i] and the historically correct thematic consonant can be used to support the hypothesis that *akin[i] was phonologically bound with some verbs further back than Proto Southeast Solomonic. With verbs where the original final consonant is reflected with *akin[i] reflexes in languages from different subgroups then it can be hypothesised that *akin[i] was indeed a suffix with these verbs before the loss of final consonants in each of these groups of languages, and thus most likely in Proto Oceanic.

For example, Proto Oceanic *taŋ ‘to cry’ is reflected with *akin[i] and a thematic consonant reflecting *s in Manam, (NNG), Gela, Kwaio and Arosi (SES), Bauan and Wayan Fijian and Samoan (Pn), as shown in Table 5. On the basis of these data a Proto Oceanic form *taŋ-akin[i] ‘to cry over, to mourn’ can be reconstructed with *akin[i] as phonologically bound to the verb stem. While it is possible that the reflexes of *akin[i] became phonologically bound after the break-up of Proto Oceanic but prior to the loss of final consonants in the Schouten languages, the Southeast Solomonic languages and the Central Pacific languages, the reconstruction of *akin[i] as phonologically dependent with *taŋ ‘to cry’ in Proto Oceanic is a plausible interpretation of these data.

---

11 The Proto Oceanic reconstruction *to[n,d]om-i- ‘to swallow’ is a simplification. Lynch (2001) shows that a number of distinct, but phonologically related, forms can be reconstructed for Proto Oceanic with the meaning ‘swallow’.

12 See Blust (1977) on the different ways in which such reduplicated Proto Austronesian verbs are reflected in Proto Oceanic.

13 The sound correspondences of the thematic consonants in Table 5 are regular. Thus Proto Oceanic *s is retained as such in Samoan and in this environment (before i) in Kwaio and Arosi, whereas in Gela it is regularly reflected as h. In Bauan and Wayan Fijian, Proto Oceanic *s is reflected as ð, and in Manam *s becomes r in medial position. Proto Oceanic *m is retained as such in Manam, Kwaio, Tongan and Samoan, and Proto Oceanic *p is reflected as h in Sa’a and Arosi, as v in Mota and Wayan Fijian and as f in Tongan.

14 The Boumaa and Wayan Fijian reflexes of *akin[i] suggest that the criteria for phonological dependence are not as straightforward as presented here. Thus, although in these languages the reflexes of *akin[i] are phonologically independent in terms of stress assignment, they are phonologically dependent in terms of occurring with thematic consonants. It is possible that the same was true in Proto Oceanic. That is, that *akin[i] was phonologically dependent in some respects, and thus thematic consonants were retained before its reflexes, but was independently stressed. Lynch (2000) presents a reconstruction of
Table 5: Reflexes of *akin[i] with the ‘correct’ thematic consonant

<table>
<thead>
<tr>
<th></th>
<th>Root form</th>
<th>Form with *-i</th>
<th>Form with *akin[i]</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘cry’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POc</td>
<td>*taniš</td>
<td>*taniš-i-</td>
<td>*taniš-akin[i]-</td>
</tr>
<tr>
<td>NNG</td>
<td>Manam</td>
<td>tajji</td>
<td>tajji-i-i</td>
</tr>
<tr>
<td>SES</td>
<td>Gela</td>
<td>taji</td>
<td>taji-hagi</td>
</tr>
<tr>
<td>SES</td>
<td>Kwaio</td>
<td>ani</td>
<td>ani-te’eni-</td>
</tr>
<tr>
<td>SES</td>
<td>Arosi</td>
<td>anji</td>
<td>anji-ta’i</td>
</tr>
<tr>
<td>Fij</td>
<td>Bauan</td>
<td>taji</td>
<td>taji-ða</td>
</tr>
<tr>
<td>Pn</td>
<td>Wayan</td>
<td>taji</td>
<td>taji-ðakini-</td>
</tr>
<tr>
<td>Pn</td>
<td>Samoan</td>
<td>taji-</td>
<td>taji-si-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>taji-sa’i</td>
</tr>
<tr>
<td>‘bury, plant’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POc</td>
<td>*tanum</td>
<td>*tanum-i-</td>
<td></td>
</tr>
<tr>
<td>NNG</td>
<td>Manam</td>
<td>tano</td>
<td>tano-m-i</td>
</tr>
<tr>
<td>SES</td>
<td>Kwaio</td>
<td>ano</td>
<td>ano-me’eni-</td>
</tr>
<tr>
<td>Pn</td>
<td>Tongan</td>
<td>tanu</td>
<td>tanu-ma’i</td>
</tr>
<tr>
<td></td>
<td>Samoan</td>
<td>tanu</td>
<td>tanu-ma’i</td>
</tr>
<tr>
<td>‘swim’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCEOc</td>
<td>*garu(p)</td>
<td>*garu-p-i</td>
<td>*garu-p-akin[i]-</td>
</tr>
<tr>
<td>SES</td>
<td>Arosi</td>
<td>—</td>
<td>’aro-ha’i</td>
</tr>
<tr>
<td>SES</td>
<td>Sa’a</td>
<td>olo, oloolo</td>
<td>olo-hi</td>
</tr>
<tr>
<td>NCV</td>
<td>Mota</td>
<td>garu</td>
<td>garu-vag</td>
</tr>
<tr>
<td>Fij</td>
<td>Wayan</td>
<td>qua</td>
<td>qua-vi</td>
</tr>
<tr>
<td>Pn</td>
<td>Tongan</td>
<td>kaukau</td>
<td>kau-faki</td>
</tr>
</tbody>
</table>

With the other two verbs in Table 5, reflexes of *akin[i] with the historically correct thematic consonant are found only in languages of the Central/Eastern Oceanic dialect complex. Thus Proto Oceanic *tanum ‘to bury, plant’ is reflected with *akin[i] and a thematic consonant reflecting *m in Kwaio (SES), Tongan and Samoan (Pn). The Proto Central/Eastern Oceanic form *garu(p) ‘to swim’ is reflected with *akin[i] and the historical thematic consonant in Arosi and Sa’a (SES), Mota (NCV) and Tongan (Pn). Thus *akin[i] is reconstructable as phonologically dependent with these two verbs before the break-up of the Central/Eastern Oceanic dialect complex.

The fact that *akin[i] was phonologically bound with different verbs at different stages within the history of Oceanic suggests that *akin[i] was being reanalysed as a phonologically dependent form on a lexeme-by-lexeme basis. Thus, in Proto Oceanic in most of its occurrences *akin[i] was phonologically independent, but in combination with *taniš ‘to cry’, and probably also a number of other verbs, it was phonologically dependent on the verb. After the break-up of Proto Oceanic, but before the break-up of the Central/Eastern Oceanic dialect complex, *akin[i] had become phonologically bound with more verbs, including *tanum ‘to bury, plant’ and *garu(p) ‘to swim’. Before the break-up of

Proto Oceanic stress, but further research would be needed to reconstruct the stress assignment of particular morphemes, such as *akin[i].
Proto Southeast Solomonic *akin[i] had become phonologically bound with even more verbs, including *to[n,d]om ‘to swallow’, *susu(p) ‘to suckle’ and *siki(l) ‘to poke out’.\textsuperscript{15} These developments are shown in Table 6.

### Table 6: Verbs reconstructable with *akin[i] as a suffix

<table>
<thead>
<tr>
<th></th>
<th>Proto Oceanic</th>
<th>Proto Central/Eastern Oceanic</th>
<th>Proto Southeast Solomonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘to cry’</td>
<td>*taŋis</td>
<td>*taŋis</td>
<td>*taŋi</td>
</tr>
<tr>
<td></td>
<td>*taŋis-i-</td>
<td>*taŋis-i-</td>
<td>*taŋi-si-</td>
</tr>
<tr>
<td></td>
<td><strong>taŋis-akin-i-</strong></td>
<td>*taŋis-akin-i-</td>
<td>*taŋi-saki-ni-</td>
</tr>
<tr>
<td>‘to bury, plant’</td>
<td>*taŋum</td>
<td>*taŋum</td>
<td>*taŋu</td>
</tr>
<tr>
<td></td>
<td>*taŋum-i-</td>
<td>*taŋum-i-</td>
<td>*taŋu-mi-</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>taŋum-akin-i-</strong></td>
<td>*taŋu-maki-ni-</td>
</tr>
<tr>
<td>‘to swim’</td>
<td></td>
<td><strong>gaRup-akin-i-</strong></td>
<td>*galu-vaki-ni-</td>
</tr>
<tr>
<td>‘to swallow’</td>
<td>*to[n,d]om</td>
<td>*to[n,d]om</td>
<td>*to[n,d]o</td>
</tr>
<tr>
<td></td>
<td>*to[n,d]om-i-</td>
<td>*to[n,d]om-i-</td>
<td>*to[n,d]o-mi-</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>to[n,d]o-maki-ni-</strong></td>
<td></td>
</tr>
<tr>
<td>‘to suckle’</td>
<td>*susu</td>
<td>*susu</td>
<td>*susu</td>
</tr>
<tr>
<td></td>
<td>*susu-</td>
<td>*susu-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*susu(p)</td>
<td>*susu(p)</td>
<td>*susu-vi</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>susu-vaki-ni-</strong></td>
<td></td>
</tr>
<tr>
<td>‘to poke out’</td>
<td></td>
<td><strong>siki-laki-ni-</strong></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{15} It is, of course, possible that the ‘correct’ thematic consonants with reflexes of *akin[i] are not historical but have been inserted by analogy with reflexes of *-i before which original word-final consonants were retained. While this is a plausible analysis of the contemporary data and probably represents the actual history of some thematic consonants with reflexes of *akin[i], the account presented here is taken to be more likely overall. Within contemporary languages it is only some verbs which occur with *-i and *akin[i] reflexes and have corresponding thematic consonants. For example, as mentioned above, in Gela only 56\% of verbs with -\textit{Ci} and -\textit{Caŋi} have corresponding thematic consonants. Thus this type of analogy is not taken to have played a major role in the history of thematic consonants with *akin[i]. For this reason it is considered more likely that corresponding thematic consonants with *-i and *akin[i] with particular verbs across languages from different subgroups reflect the retention of an original word-final consonant rather than a process of analogy with the same verb in several different languages or groups of languages.
4 The grammatical status of *akin[i]

*Akin[i]* is widely reflected in contemporary Oceanic languages as both a transitivising element within the verb complex and a preposition. Figure 1 schematises these two types of structures. The structures under (i) represent *akin[i]* as a preposition with intransitive and transitive verbs, a use that is reflected in Tamambo and Woleaian. In such constructions the preposition, a reflex of *akin[i]*, occurs with an object marker that indexes the participant denoted by the object of the preposition. The structure in (ii) represents the kind of structure found with *akin[i]* reflexes in languages like Manam and Boumaa Fijian, where *akin[i]* is part of the verb complex and the participant it introduces is expressed as the object argument.

\[
\begin{array}{ll}
(i) & \text{[S=V]} \text{VC} \\
 & \text{[S=V(-i)=OBJ]} \text{VC} \\
 & \text{[akin-i=OBJ NP]} \text{NP} \\
 & \text{[akin-i=OBJ NP]} \text{PP} \\
(ii) & \text{[S=V *akin-i=OBJ]} \text{VC} \\
 & \text{[O]} \text{NP} \\
 & \text{[akin-i=OBJ NP]} \text{NP} \\
\end{array}
\]

**Figure 1:** Constructions with *akin[i]* reflexes in contemporary Oceanic languages

The historical correspondence of these constructions is apparent from the cognate morphology, that is, the presence of a morpheme reflecting Proto Oceanic *akin[i]*, and the cognate function of *akin[i]*, allowing the expression of participants with certain semantic roles with particular semantic classes of verbs (see Table 1). However, the grammatical status of *akin[i]* in Proto Oceanic is not so apparent. The two types of construction in Figure 1 suggest that in Proto Oceanic *akin[i]* was:

(a) the head of a prepositional phrase;

(b) a post-verbal element within the verb complex; or

(c) analysed variably by speakers as either (a) or (b).

However, while it is clear that the string of elements VERB AKIN-I=OBJECT MARKER NOUN PHRASE was present in Proto Oceanic, it is not immediately apparent which scenario is the most fitting reconstruction and if either of the two structural analyses presented in Figure 2 can be shown to be an earlier one. A change from one structure to the other, in either direction, through syntactic reanalysis would be equally plausible.\(^{17}\) Harris and Campbell (1995:72) establish that a necessary condition of reanalysis is ‘that a subset of

---

\(^{16}\) Figure 1 makes certain generalisations about the structure of the verb complex in Oceanic languages. Most Oceanic languages have preverbal subject markers that are here represented as proclitics, which is only true for some languages. Markers of tense, aspect and mood also tend to occur preverbally in Oceanic languages, but are not indicated here. Subject noun phrases are also not indicated. In languages which are SVO the subject noun phrase would precede the verb complex. In other languages, with VSO word order the subject noun phrase would occur between the verb complex and the object noun phrase.

\(^{17}\) See for example, Durie’s (1988) study of verbal prepositions in Oceanic languages, which demonstrates processes of change by which a verb within a serial verb construction can: (a) become more strongly associated with the other verb(s) in the construction (the centripetal tendency); or (b) become more strongly associated with a peripheral constituent and become an adposition (the centrifugal tendency).
the tokens of a particular constructional type must be open to the possibility of multiple structural analyses, where one potential analysis is the old one (applicable to all tokens) and the other potential analysis is the new one (applicable to a subset). That both of the structural analyses of constructions with *akin[i] in Figure 2 were potentially available to Proto Oceanic speakers is supported by both being reflected in contemporary Oceanic languages (Figure 1).

(i) \([S/A=V]_{VC} \ [akin-i=OBJ \ NP_X]_{PP}\)
(ii) \([S/A=V \ akin-i=OBJ]_{VC} \ [O_X]_{NP}\)

**Figure 2:** Two analyses of clauses with *akin[i]

In terms of syntactic reconstruction, Harris and Campbell (1995:353–360) advocate the use of archaisms and morphology as aids for determining the relative chronology of particular structural analyses. Morphology as ‘concrete and phonologically endowed’ lends itself more readily to reconstruction with traditional methods, and its close relationship to syntax means that morphological reconstruction is often indicative of aspects of proto-syntax (Harris and Campbell 1995:358).

There is morphological evidence to suggest that Proto Oceanic *akin[i] was originally part of the verb complex. Non-Oceanic cognates of Proto Oceanic *akin[i] reflect an apparent form *akən, the expected reflex of which would be **akon in Proto Oceanic. The final *-i of *akin[i], as proposed by Clark (1973), is best analysed as historically reflecting the Proto Oceanic transitive suffix *-i. The form of the medial vowel as *-i- rather than the expected **-o- appears to result from a sporadic change of regressive vowel assimilation. The transitive suffix *-i in Proto Oceanic, as in many contemporary Oceanic languages, most likely occurred preceding the object enclitics and phonologically bound to the final element (verb or post-verbal modifier) of the verb complex (Evans 2003:117–118). Thus, that *-i occurred with *akin[i] is evidence that it was a post-verbal element in the verb complex. The occurrence of *-i with the antecedent of *akin[i], and the subsequent vowel assimilation change, must have occurred prior to the break-up of Proto Oceanic as only *akin[i] with a high front medial vowel and not **akon is reflected in Oceanic languages.

The reconstruction of *taŋis-akin[i] with the specialised meaning ‘to mourn’ also suggests that Proto Oceanic speakers analysed *akin[i] and the preceding verb stem as a unit, at least with some verbs. As demonstrated by Bybee (1985), the linguistic expression of two or more meaning elements is in part dependent on their relevance to each other. Bybee (1985:13) defines relevance as the semantic content of one meaning element.

---

18 The cognacy of the Oceanic forms reflecting *akin[i] and non-Oceanic forms reflecting *akən is supported not only by the similarity in form but also by the similarity in function and distribution (Evans 2003).

19 The transitive suffix *-i is also reflected outside of the verb complex with verbal prepositions in a number of Oceanic languages. As a number of verbal prepositions have been shown to derive from lexical verbs (see Pawley 1973; Lichtenberk 1985; Durie 1988; Ross 1988), *-i in Proto Oceanic is taken to have occurred primarily, if not solely, within the verb complex.
directly affecting or modifying the semantic content of another. However, it is not only the semantic features of the two elements which determine their relevance to each other but also the cultural or cognitive salience of their combination. Bybee’s (1985) cross-linguistic study predicts that if two meaning elements are highly relevant to each other they are more likely to have inflectional or lexical expression. For Pawley and Syder (1983:209) a lexicalised sequence is not only characterised by the relevance of the two meaning elements but also by a tendency to behave as a unit syntactically. The salience of the combination of *taŋis and *akin[i] to Proto Oceanic speakers can be seen from its specialised meaning, which I would argue also supports the hypothesis that, at least in this instance, *akin[i] formed a conceptual, and most likely also a syntactic, unit with the preceding verb stem.

While these two pieces of evidence suggest that speakers of Proto Oceanic did analyse *akin[i] as part of the verb complex, that is, structure (ii) in Figure 2, there is still the question of whether structure (i), *akin[i] as a preposition, was also an analysis made by Proto Oceanic speakers. One type of evidence that is indicative of multiple analyses in the history of particular morphemes is the presence of multiple reflexes within one or more daughter languages (Harris and Campbell 1995:81–83).

In a number of Oceanic languages there are two reflexes of *akin[i], one which is within the verb complex and one which is a preposition. For example, alongside the preposition yagili, Woleaian also has a verbal suffix -agi.20 Other languages with two reflexes of *akin[i] include, Lengo (SES), Boumaa and Wayan Fijian, and Samoan and Tongan (Pn). This implies that in the history of each of these languages *akin[i] had at least two structural analyses (those schematised in Figure 2). While it is possible that the prepositional reflexes of *akin[i] are the result of parallel development, it seems likely that the dual reflexes of *akin[i] in such a wide range of languages reflects the presence of the two structural analyses of *akin[i] in their common ancestor, in this case Proto Central/Eastern Oceanic. In Western Oceanic languages most reflexes of *akin[i] are transitivising suffixes within the verb complex, although there are occasional reflexes which do have uses such as Teop (MM) ki (Mosel and Thiesen 2007). While it thus seems likely that in Proto Oceanic there were also two structural analyses of *akin[i], parallel development cannot be easily excluded.

A question which remains is what the original grammatical status of *akin[i] was. Harrison (1982), and also Lichtenberk (2005), suggest that *akin[i] may have been a lexical verb used in serial verb constructions. Crowley (2002) demonstrates that similar types of serial verb constructions are widely distributed in contemporary Oceanic languages, but notes that it is difficult to exclude the possibility that such constructions arose through contact with non-Austronesian languages and spread as the result of diffusion. He (2002:168) concludes, however, that serial verb constructions must represent ‘a fairly old pattern’ within Oceanic which goes back at least to the ‘early offshoots of Proto Oceanic’. Thus, it is not implausible to reconstruct the occurrence of *akin[i] within a serial verb construction early in its history within Oceanic. However, demonstrating a verbal origin of *akin[i] is difficult since, to the best of my knowledge, *akin[i] is not

---
20 The function of Woleaian -agi is somewhat different to the transitivising suffixes that reflect *akin[i] in other Oceanic languages (see Sohn 1975:131–134) and it has obviously undergone certain functional changes.
reflected as a lexical verb in any Oceanic language. For this reason the reconstruction of *akin[i] presented here will not be more precise than to state that in Proto Oceanic *akin[i] was originally a lexical verb or verbal modifier within the verb complex.

5 Concluding remarks

An examination of the different processes of change in the history of Proto Oceanic *akin[i] and its reflexes leads to a more detailed reconstruction than the seemingly straightforward process initially suggested by examples (1) to (6). Proto Oceanic *akin[i] was a morpheme which occurred in the string VERB AKIN-I=OBJECT MARKER NOUN PHRASE. It is proposed here that *akin[i] was in the midst of both phonological and grammatical change. Phonologically *akin[i] can be reconstructed as dependent on the preceding verb stem with some, but not all, lexemes. After the break-up of Proto Oceanic *akin[i] appears to have become phonologically bound with an increasing number of lexemes, probably on a lexeme-by-lexeme basis.

Grammatically there was also variation in the analysis of *akin[i] in Proto Oceanic. Its occurrence with the transitive suffix *-i indicates its original presence within the verb complex, and thus Proto Oceanic speakers appear to have analysed *akin[i] as a verb or verbal modifier which allowed the expression of an object argument with one of a set of semantic roles (see Table 1). It is likely that at least some speakers of Proto Oceanic had a second structural analysis of *akin[i], namely as a preposition introducing an oblique argument. After the break-up of Proto Oceanic this innovative analysis appears to have become more frequent, taking over completely from the original analysis in some languages.

This account of the history of *akin[i] reconstructs both phonological and grammatical variation for Proto Oceanic. The present study builds on previous work on *akin[i] (Pawley 1973; Harrison 1982), but by considering the history of the morpheme in terms of likely processes of change provides a more detailed reconstruction which better accounts for the apparently conflicting data from contemporary reflexes of *akin[i].

References


In Tukang Besi, a non-Oceanic Austronesian language of Indonesia, the cognate of *akin[i], ako occurs as an applicative suffix and a free form which can be described as an atypical verb. Ako does occur as a main verb, but only with a benefactive meaning (Donohue 1999). While the Tukang Besi data is crucial in examining the origins of Proto Oceanic *akin[i], on its own it is not enough to reconstruct a verbal origin of *akin[i] in pre-Proto Oceanic.


The extinct language of Vunaqumu

PAUL GERAGHTY

Introduction

Fiji is not a place that is usually associated with dead or dying languages, and indeed language death is not occurring there at anything like the rate reported for parts of Melanesia and Australia. Nevertheless, a number of the 300 or so communalects are in the process of disappearing, and probably all are gradually losing their distinctiveness under the influence of Standard Fijian and regional prestige languages, such as Cakaudrove and Nadroga (Mangubhai and Mugler 2006:89–90). This paper will focus in particular on the dead communalect of Vunaqumu.¹

The Vunaqumu people

In the early 1980s, I trekked up to the headwaters of the Wainimala River in central Vitilevu, in what is now the province of Naitasiri (formerly Colo East), with the intention of doing research on the Emalu language (also known as Noemalu, Noemalo or Noimalo). The people who now reside there are known to have come over fairly recently — perhaps two hundred years ago — from the headwaters of the Sigatoka River to the west, in the vicinity of Nasauvakarua, so, unlike most other residents of the Wainimala, they speak a Western Fijian language. In the course of my research, which lasted approximately five days, the people told me of another language that used to be spoken in the vicinity, but had been displaced by the language of the invaders from the west, and was now no longer

¹ It gives me great pleasure indeed to contribute this paper to a collection to honour the life and work of Andrew Pawley. He was the first Oceanic linguist I ever met — in Auckland in July 1970, after I had spent my first year in Fiji and decided I would like to study the languages there — and has given me encouragement and constructive criticism ever since, particularly as my PhD supervisor at the University of Hawai‘i, even putting up with me as a house guest for a few months in Auckland in 1977. May his prodigious output long continue. Mo cola balavu riki na mōmō leva!

This is a revised and updated version of a paper I first presented at the Fifth International Conference on Oceanic Linguistics, at The Australian National University, Canberra, 14–16 January 2002, entitled ‘The language of Vunaqumu and other extinct or moribund Fijian communalects’. The research was partly funded by the Unesco Project for Indigenous Language Revitalisation and Preservation.
spoken. They called it the Vunaqumu\textsuperscript{2} language, because it was the original language of the Vunaqumu people.

The Vunaqumu people are reported (de Marzan 1913) to have their origins in the region of Nakauvadra, just inland of the northern coast of Vitilevu, a place which many tribes of Vitilevu now claim to have originated from (Gifford 1952:340). Indeed, one of the early chiefs of Vunaqumu was called Degei (Brewster 1890), the name of the supreme deity of Nakauvadra and much of eastern Fiji, while an early chief of the people of Nawacakena (or Waicakena), close relations of the Vunaqumu people, was called Vukinavanua (de Marzan 1913), literally ‘the land turns’, an alternative name for Degei.

The Vunaqumu people moved from there southwards to the headwaters of the Nailuva River, in what is now the south-west part of Ra province, living there with their close relations the people of Nawacakena. Then a dispute arose between the two tribes, and the Vunaqumu people, under the leadership of Delaivunaqumu, left to journey westwards. Brewster (1890) states that the chief of Vunaqumu in 1884, Ro Vunimatana, was the eighth in line after Delaivunaqumu, so by allowing 25–30 years per generation, we can estimate that this incident took place some time in the second half of the 17\textsuperscript{th} century. They settled briefly in Naleka, between Nadrau and Navai, then followed the Wailoa downriver, ascended the Wainimala, and finally settled at Nakurukuruvakatini, a hill near the confluence with the Wainamu, in the vicinity of which the senior branch of the tribe remains to this day (de Marzan 1913), principally in the villages of Korovou and Nasauvere and the more recent settlement of Tubarua.

Soon after the settlement of Nakurukuruvakatini, the Naitaubale branch of the tribe crossed the watershed to the south and moved to the upper Wainikoroiluva River, where they remain to this day in the villages of Naqwarai and Naraiyawa in what is now Namosi province. Although their language is distinct in some minor respects from that spoken by their Namosi neighbours, they retain very little of the Vunaqumu language. Another branch, the Naitavuni, moved south-east to the Waidina River, between the Namosi and Waimaramo people, where they now occupy the villages of Naseuvou and Delailasakau. They also have little memory of the Vunaqumu language. Other groups are said to have ventured as far as Suva, Waisomo in Beqa, and Matuku in Lau.

Approximately two hundred years ago, Emalu people began to settle in what was Vunaqumu territory in the upper Wainimala, and the Vunaqumu people remaining there have now completely assimilated to the Emalu language. When I was conducting fieldwork there in the 1980s, I was told that the last fluent speaker of Vunaqumu was a lady who died around 1933, and that only a handful of elderly people had heard Vunaqumu spoken when they were children. It is therefore likely that the generation who had even passive knowledge of Vunaqumu have by now all passed away.

Sentences

The following sentences were said by informants to have been uttered by fluent speakers of Vunaqumu:

\begin{itemize}
\item \textit{Na yava xo mani vaxa mai qoinixa?} ‘What are you doing here?’
\item \textit{Na yava xo xwā jixo?} ‘What are you saying?’
\item \textit{Vaxoitā vai?} ‘How is it? How is it done?’
\end{itemize}

\textsuperscript{2} Note that in Fijian orthography, \textit{q} represents a prenasalised voiced velar stop \([ŋg]\), \textit{g} a velar nasal \([ŋ]\).
Na naura xa dani ruxuta na vā. ‘The prawn is found in rocks’
(lit. ‘habitually enters the rocks’).

Na yava na yana qō? ‘What is the name of this?’

Au xauta mai au mai solia vai xexo. ‘I brought it here and gave it to you.’

Vō drī jixo mai xē? Mai lagi. ‘Where have you come from? From above/upriver.’

Do xauta ra sūsū do mani taxivaxa mai wai. ‘Take it children and go and draw water with it from the river.’

E dua na wā mo taberaxa jixo i domodomonilagamu, dreta mai na qito, me ra siri na yalewa. ‘You should tie (?) a vine to your ankle and pull the firewood so the women can pass.’

Voi bogi tū mai na vanua. ‘It’s getting dark.’

Xau voi rai xoto mai noqu bure. ‘I was casually looking out from my men’s house.’

Voi xwā me gwani e dua me drī tani. ‘He/she said that no-one should go away.’

Au tari sa boi nova na veli. ‘I haven’t yet seen a forest gnome.’

Xau drī jixo ma xi rā. ‘I am coming down/downriver.’

Xo yei xa borota na lagana na sūsū nāqō? ‘Who slashed (?) this child’s leg?’

Au nova na noqu vuzi, so qoi sexēria tū nīgwā. ‘I looked at my banana tree and it was just starting to shed the leaves of its inflorescence.’

E drasavaxa jixo na loma na magaqu. ‘The inside of my vagina is red and inflamed’
(i.e. I am very angry).

Vocabulary

The following is a vocabulary of Vunaqumu words, excluding pronouns and demonstratives, which will be listed in the grammatical notes below. Standard Fijian (SF) equivalents are given in brackets after the definition, unless the SF word is phonemically the same.

A hyphen after a noun indicates it is suffix possessed; with verbs, a hyphen separates the base from the consonant of the transitive suffix, or -tx, which represents the long (oblique) transitive suffix -taxi/taxa. If a verb ends in a consonant, it is obligatorily transitive, being suffixed with -i or -a. The instrumental prefix i- is ignored in alphabetisation.

Abbreviations: ADV adverb, ART article, CNJ conjunction, LN locative noun, N noun, PO postverbal particle, POSS.CLF possessive classifier, PRE preverbal particle, PREP preposition, TR transitive, V verb, V.ACT active-type verb, V.PASS passive-type verb.

-a  transitive suffix + 3s object pronoun
-bī  N bamboo (bitu)
-axa  oblique transitive suffix + 3s object pronoun
-bibī  V heavy
-axi  oblique transitive suffix
-bogi  N night
-bajī-  N tooth
-boī  PRE even, a little (bau)
-balabala  N tree-fern
-boro-t  V.ACT slash (?)
-balava  V long, tall
-bote  N taro leaves as food (rourou)
-bale  PO down (sobu)
bou  N kingpost (main housepost extending to ridgepole)
boxa N taro (= qau) (dalo)
bury N men’s house
buta V cooked
buyobuyo V white (yulavula)
carī-tx V.ACT lie to (lasu-tx)
cauravou N young man
cebe-t V.PASS slice
icoi N (meal) non-starch part
icoxa N house-beam
dabe V sit
dai N trap
dani PRE habitually, always (dau)
davuxi N pit (davuke)
daxana CNJ lest, in case (dē)
daxi-tx V.PASS put down (biu-t)
domodomonilaga- N ankle
(domodomoniyava-)
dovu N sugarcane
idoxo N digging-stick (isau)
dua V one
duatanji V different, strange
dui PRE each
duru N house-post
duvu N bottle (tavaya)
drasavaxa V red and inflamed (wādamu)
drau N thatch
dre-t V.ACT pull
drex-tx V.ACT carry (burden) on back
dreu V ripe
dri V go (lako)
gwani V do not, should not (kua ni)
i PREP in, at, on (e)
-i transitive suffix
ibe N pandanus for mats (voivoi)
ibe N mat
jixo V stay
jixo PO continuous tense
laga- N leg, foot (yava-)
lagi LN above, upriver (cake)
lave-t V.ACT,V.PASS lift
lemu V big (levu)
lololo V black (loaloa)
loma- N inside
lutu V fall
ma PO hither (=mai)
maduxi V (wood, fruit) dry (madū)
maga- N vagina
mai PO hither (=ma)
mala N yam slice (de Marzan only, not confirmed) (kome)
mamau V sated, full after eating
mani PRE go and (lai)
masa V empty of liquid (maca)
mata- N eye
matanadarava N doorway
(matanaxwātuka) (kātuka)
matanaxwātuka N doorway
(matanadarava) (kātuka)
matāsawa N beach
mataxwā N morning (mataka)
mate V die
matua V mature
me CNJ irrealis
me- POSS.CLF drink, possessive marker
mila-Ø V.ACT scratch (itch)
mo CNJ+PRONOUN that you (sing)
muri-Ø V.ACT follow
musu-x V.PASS break
na ART the, a
na PREP of
na- N mother (tina-)
namo N pile of yams (duqele)
namu N yam (de Marzan only, not confirmed, perhaps confused with the above)
nanum V remember
naura N prawn (ura)
naxwā V good (vinaka)
ne- POSS.CLF general possessive marker before /i/
niavi ADV yesterday (nanoa)
nigwā PO only, focus marker (gā)
no- POSS.CLF general possessive marker
nov V see (rai-c)
obo V clap with hollowed hands
oji V finished
qau N taro (= boxa) (dalo)
qito N firewood (buka)
The extinct language of Vunaqumu

qoi PRE then, next (qai)
qwara N hole
ra non-singular vocative prefix
rā LN below
rā PO please, first (mada)
raut V enough
rou V dive for prawns (nunu ura)
ruxu-t V.ACT enter (curu-m)
sa PRE continuing aspect (se)
sala N pig track (calala)
sālemu N road (gaunisala)
sara PO straightaway, very
isāsari N rib (saresare-)
isavu N what one is known for (icavu)
sesēria V (banana) shed the leafy covers of the inflorescence (?)
seu V run hard (=seuwai) (sorova)
soriji N reeds bound together in wall construction
soro-tx V.ACT hurry (totolo)
suxu- N hand, arm (liga-)
sūsū N child (gone)
tabex V tie(?)
tagwane N male (tagane)
tai- N younger same-sex sibling (taci-)
tama- N father
tani PO away, different
tara-Ø V.PASS build (house)
tari PRE not (sega ni)
taur V take
tavi V cut leaves (pandanus)
tavo V pull up (grass for thatch) (cavu)
taxi-v V.ACT draw water with
teitei V plant
tōā N fowl, hen (toa)
tū V stand
tuba LN outside
tuxun V say, tell
uso V call (kaci)
va PREP at, in before na (e)
vā N rock, stone (vatu)
vai LN where (vei)
vai- reciprocal prefix, general human object prefix (vei-)
vale N house
evaua N land
vaxa V do (=yaxa-V) (caka-V)
vaxarogo V hear (rogo-c)
vaxasaqa V cook
vaxasasa V hunt pigs
vaxatata V speak, language (vosa)
veli N forest gnome
vili-x V.PASS pick (fallen fruit)
visa V how many, how much (vica)
vō = vai + xo
vōi PRE inchoative aspect (sā)
vua N yam (uvī)
vūā PO to him/her
vudi N plantain
vusi N swamp (vuci)
vuta V refuse (bese)
wā N vine
wacaxa- N stomach (kete-)
wai N water
wai N, LN river
wāwā V wait
warā V no, not exist (sega)
were N yam garden
xaba PO up (cake)
xai PREP and (kei)
xaisāsā N pole, rafter (kau, isā)
xālava V stride
xalou N spirit, god
xanaga N food (kākana)
xatu N fathom
xau N wood, tree
xau V.PASS bring, take
xe- POSS,CLF eat possessive marker
xē in it + (=? xina) (kina)
xi PREP to (non-human) (i)
xina PO in it, on it, there + (=? xē)
Grammatical notes

The phonology of Vunaqumu is identical to that of Nadrau (Geraghty 1983:28; Kikusawa 2001), and all other languages of the upper Wainimala including Emalu, with /t/ and /d/ palatalised (written as j and z respectively) before a high front vowel, and /k/ fricativised (written x). However, unlike most other Fijian communautics other than Lau, there is no diphthongisation, so that penultimate high vowels retain primary stress even after low or mid vowels, e.g. xaú-ta ‘bring it’, taí-qu ‘my younger same-sex sibling’.

Possession is also, as far as can be determined, identical to that of Nadrau (Geraghty 1983:235, 247), with a lack of the portmanteau POSS.CLF+i forms used before proper nouns elsewhere in Eastern Fiji, e.g. nona vale xo Jone ‘Jone’s house’ (Standard Fijian na vale nei Jone).

The pronouns also closely resemble Nadrau (gaps in the table mean no data available):

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>Preverbal</th>
<th>Suffix possessive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td>yau</td>
<td>xau,au</td>
<td>qu</td>
</tr>
<tr>
<td>2s</td>
<td>xexo</td>
<td>xo</td>
<td>mu</td>
</tr>
<tr>
<td>3s</td>
<td>xea</td>
<td>e, xa</td>
<td>nna</td>
</tr>
<tr>
<td>1de</td>
<td>xairau</td>
<td></td>
<td>irau</td>
</tr>
<tr>
<td>1te</td>
<td>xai</td>
<td></td>
<td>itou</td>
</tr>
<tr>
<td>1pe</td>
<td>xaimamu</td>
<td></td>
<td>imamu</td>
</tr>
<tr>
<td>1di</td>
<td>xēdaru</td>
<td>daru</td>
<td>daru</td>
</tr>
<tr>
<td>1ti</td>
<td>xedatou</td>
<td>to</td>
<td>da</td>
</tr>
<tr>
<td>1pi</td>
<td>xeda</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2d</td>
<td>xemudrau</td>
<td>drau</td>
<td></td>
</tr>
<tr>
<td>2p</td>
<td>xamū</td>
<td>mū</td>
<td></td>
</tr>
<tr>
<td>3d</td>
<td>ratou</td>
<td>ra</td>
<td>dra</td>
</tr>
<tr>
<td>3t</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The extinct language of Vunaqumu

The demonstratives however show varied affinities. In the first person, they are similar to the Beqa region (the nearest part of which to Vunaqumu is Deuba on the south coast of Vitilevu, near Navua), and the second person is identical to Nadrau and numerous other eastern communalects, including some from Lau. The third person is unique, but the manner of its formation, with $nā-$ prefixed to the first person form, is identical to that of Emalu and related Western Fijian communalects:

Table 2: Demonstratives

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>Postnominal</th>
<th>Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>iqō</td>
<td>qō</td>
<td>qoinixa</td>
</tr>
<tr>
<td>II</td>
<td>yā</td>
<td>yā</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>$nā$qō</td>
<td>$nā$qō</td>
<td></td>
</tr>
</tbody>
</table>

The most intriguing grammatical morpheme of Vunaqumu is the transitive suffix, which seems totally at odds with its geographical location. In the eastern half of Vitilevu, the transitive suffix is -$ia$ in the north and -$e$ in the south, with -$a$ being found only in isolated coastal pockets: Deuba in the south, the Bau-Rewa area in the south-east, and the Verata area in the east and north-east (Geraghty 1983:262). Finding the -$a$ suffix in the middle of the highlands of Vitilevu is therefore quite unexpected. The most plausible historical explanation is that the -$a$ suffix area was previously far more extensive than it is now, and that the languages of south-east Vitilevu with -$e$ suffixation are intrusive. This theory has some support in oral tradition. For example, traditions of both Namosi and Deuba agree that the Deuba people formerly occupied at least part of what are now Namosi lands, the Namosi people having come fairly recently from the Taivugalei area in the eastern highlands of Vitilevu. Therefore, in this respect, Vunaqumu is a relic area.

Lexical affinities

The following lexical items appear to be unique to Vunaqumu: $bī$ ‘bamboo’ (cf. SF $bitu$), $cari$-$tx$ ‘lie to’, $dani$ ‘habitually, always’, $drī$ ‘go’ (cf. SF $drī$ ‘rebound’), $lagi$ ‘above, upriver’ (cf. Proto Central Pacific (PCP) *lagi ‘sky’), $lemu$ ‘big’ (cf. SF $levu$), $naura$ ‘prawn’ (apparently coalescence of article $na$ with PCP *ura), $nov$ ‘see’, $sūsū$ ‘child’, $uso$ ‘call’, $vā$ ‘rock, stone’ (cf. PCP *vatu), $vaxatata$ ‘speak language’ (cf. widespread Western Fijian $tata$). Both $bī$ and $vā$ appear to show irregular loss of final -$tu$.

The following items appear to be shared with Emalu: $gwani$ ‘do not, should not’ (also Navosa), $nigwā$ ‘only, focus marker’, $voi$ ‘inchoative aspect’ (same function as SF $sā$).

The following items appear to be shared with Nadrau (and/or its close relative Nabobuco): $ibe$ ‘pandanus for mats’ (cf. SF $ibe$ ‘mat’), $na$ ‘of’ (also much of Nadroga and Navosa), $na$-$a$ ‘mother’ (also Wainibuka), $somu$ ‘drink’ (also much of western Fiji and Rakiraki), $suxu$-$a$ ‘hand, arm’, $vua$ ‘yam’ (also much of south-eastern Vitilevu), $wacaxa$-$a$ ‘stomach’ (also parts of eastern Vitilevu), $xaba$ ‘up’ (also parts of eastern Vitilevu).

Finally, $lagi$-$a$ ‘leg, foot’ is shared with Kadavu, and the preverbal negativiser $tari$ ‘not’ with parts of Rewa.
Historical phonology

Whereas the Proto Oceanic and Proto Central Pacific glottal stop is usually lost in Fijian languages, there is a small number of words in some communalecots of Vitilevu where it is reflected as /k/ (Lynch 2003:237). Two words in Vunaqumu also appear to reflect glottal stop as /k/: maduki ‘(wood, fruit) dry’ (cf. SF madū, Rotuma mafu ‘(coconut) ripe and hard, (wood) mature and hard’, Proto Polynesian *matu?u ‘dry (e.g. mature coconut))3 and daxi ‘put down’ (elsewhere in western and eastern Fiji dai, no known external cognates).

As already pointed out (Geraghty 1986:297–303), Vunaqumu is one of only four Fiji communalects that distinguish PCP *z from *c, a distinction which arguably goes back to the distinction between PAn *j/z/Z and *s/c, respectively. Indeed, an additional Vunaqumu form not considered in Geraghty (1986), yata ‘bad’ < PAn *zaqat, reduces the number of irregular developments to be explained, and thus adds weight to the hypothesis of the retention of a distinction between two palatals at Proto Oceanic level.

Conclusion

I make no apology at all for offering a data-rich paper in this volume in honour of Andrew Pawley — indeed I think it is highly appropriate, in view of Andrew’s own life-long interest in field-work and compiling dictionaries and grammars (not to mention his chiding me on numerous occasions for ‘sitting on my data’ — so here you are, at least it’s a start!). I also would like to offer a thought on the importance of recording dead and dying languages: had I not stumbled upon the language of Vunaqumu when I was looking for something else over twenty years ago, the chances are that the relic area that is Vunaqumu would never have been recorded, and it would have been assumed that the -a transitive suffix is confined to the coast of Vitilevu. Given that Vunaqumu seems to have died a death due entirely to indigenous causes — not related to the spread of English or Standard Fijian — one wonders how many other relic areas have been obliterated in the 3000 year linguistic history of Fiji.

References

Brewster, A.B., 1890, Notes on the hill tribes of Viti Levu. MS, Fiji Government Archives.
De Marzan, P.J., 1913, Histoire de la tribu de Vunaqumu (Viti Levu, Fiji). Anthropos 8:880.

3 Note however that the source of maduki is not *matu?a (as stated in Lynch 2003:237) but *ma(d,t)u?(i,u).

The outer eastern islands of the Solomons: a puzzle for the holistic approach to the anthropology of history

†ROGER C. GREEN

Epigraph
The first phase of fieldwork in the Southeast Solomons Islands Culture History project took place between 1970 and 1972. The southeast Solomons is a complex region linguistically, and at that time was virtually unknown archaeologically. The distribution of languages, as understood in the late 1960s, provided a major part of the theoretical underpinning for the program (Davidson 1999:843).

The second phase of the Southeast Solomons project ... during the late 1970s ... showed that the linguistic evidence on which the theoretical framework had originally been based was inadequate to sustain it. Nonetheless, the project generated a considerable amount of data that continues to be analysed and published (Davidson 1999:845).

These analyses, publications, recent and ongoing research have been summarised by Green and Yen (2009), the principal investigators of the South-east Solomons project.

1 Background
The following essay provides background to and the outline for a scenario that may one day unravel the intricate relationships between language, archaeology and human biology among the various populations living in the Outer Eastern Islands of the Solomons in the period from 1595 AD to the present. It is also a tribute to a long-standing colleague — Andrew Pawley — who has shared many similar endeavours with me over four decades.1

Andrew Pawley and I began our collaboration in Pacific historical linguistics and archaeology in 1966 with two papers (Green 1966; Pawley 1966), which were subsequently

1 Beverley Lang and Ryan Lang acted as both researchers and typists for various versions of the grey literature assembled under my direction over several years; these major databases now fill a series of file boxes and were essential in writing this paper. Peter Quin, Department of Anthropology Illustrator, prepared a series of associated maps and diagrams including the one used here. Without the critical and editorial input of Valerie Green during the past six weeks this paper could not have been completed.
bundled together as a single reprint available from The Polynesian Society. They appeared under the ever-critical eye of our common mentor in the field of Pacific linguistics, Bruce Biggs. In this role it was he who supplied much useful commentary on the initial manuscript drafts.

Andrew was one of Bruce’s students from his undergraduate years through to his PhD, however much he strayed from following paths that Bruce suggested (Hollyman and Pawley 1981:14, 19; Pawley 2001:5, 2003:43–44) during that period. In fact, in later years, when truly exasperated by one of Andy’s recent heresies, Bruce commented with characteristic vigour: ‘You’ve forgotten everything I ever taught you!’ (Pawley 2002:10).

In contrast, while primarily an archaeologist, my studies in anthropology also included courses in descriptive linguistics as an undergraduate taught by Stanley Newman, and in general linguistics at the graduate level taught by Dell Hymes. In a complementary fashion, Andrew had participated in several of Jack Golson’s early archaeological excavations (Golson 2004:38) as an undergraduate, and had taken academic papers in the subject with him (Pawley 1991:10). He also attended sessions in Archaeological Method and Theory, to which I was introducing his fellow student the New Zealand archaeologist Janet Davidson (Green 2007) as part of her MA programme. Thus, Andrew’s early interest in Pacific archaeology has never flagged, and has expanded into a considerable knowledge of that field, although sensibly he has left the digging part to others. It is also the case that our combined enthusiasm for the historical aspect of linguistics, and its application to issues in the culture history of the Pacific, served as a stimulus for Bruce’s own increasing involvement in that particular subdiscipline within the very much broader areas of his research usually embraced under the rubric of linguistic enquiry (Pawley 2003:38–40).

The growing passion of Andrew and myself, along with a few others, for a more holistic approach to this topic, did bring forth stern admonitions, first from Alexander Spoehr, and then — quoting Spoehr with approval — from Bruce. At the Pacific Science Congress held in Tokyo in 1966 Alexander Spoehr sounded the following caution:

> We must exercise care not to draw conclusions in one discipline from data of a logically different order derived from other disciplines. Thus the problem of linguistic subgrouping in Oceania will be solved on the evidence of language, not archaeology. Contrariwise, the prehistorian faces certain questions of time sequence and spatial relations for which only archaeological data will suffice. The same point holds for ethno-botany and physical anthropology. This is not to say that these fields should be kept isolated. Together they can make a contribution to Oceanic prehistory greater than any can achieve separately. But the logical character of the data and the specific problems with which each discipline deals must be kept in mind when one approaches questions where the findings of all these disciplines are relevant and important. (Spoehr 1968:174)

Biggs (1972:143) commented on this admonition as follows:

> At the time I remember feeling vaguely resentful of these remarks, which appeared to me to be aimed at serious attempts at synthesis which had recently been made, including papers read at the Congress. I have had cause to think of Spoehr’s comments several times since, with approval, not resentment, and I believe his caution to be an important and timely one. We are not being told by him simply to stick to our respective lasts. If we did it might turn out that the total problem of what happened in Oceanic prehistory might not be tackled at all. Synthesis is surely desirable and essential. We are being told, however, that if we want to use the other fellow’s last we had better know exactly what it is good for.
At the interpersonal level, Bruce warned Andrew and me of the need to always keep the various strands of evidence sufficiently separate that readers could evaluate the worth of each line of evidence from one of the subdisciplines within anthropology, before engaging in any exercise that attempted to integrate them. Otherwise undue risk arises through overstepping along the lines set out in Irwin’s observations on the issues involved:

(a) The need to exercise caution when seemingly conventional ideas in voyaging theory and the normal details from archaeological settlement evidence fit well, or as Irwin (1992:205) cogently put it: ‘argument and evidence were sitting compatibly in one another’s pockets’;

(b) The need to be mindful of uncertain outcomes whenever ‘archaeology and historical linguistics have been of mutual assistance in tracking prehistoric settlement, [because] the expectations of one have sometimes unnecessarily restricted the other, and the biological data are too few to produce detailed patterns’ (Irwin 1992:2).

Not surprisingly, in various collaborative efforts since that time (Green and Pawley 1999; Pawley and Green 1973, 1984), these viewpoints became some of the basic strategies common to the Pawley/Green repertoire. Moreover, in a review of these strategies in the Pacific field, Pawley and Ross (1993:452) suggest that while biological anthropology lagged behind in such integrative endeavours, it too would become a prominent player. That it now has become a respected contributor is exemplified by the recent volume edited by Friedlaender (2007), which includes historical linguistics, archaeology and molecular-based human biology. In summary, during Andrew’s long innings at this particular crease, he has established ample precedents for the use of these strategies. I see no reason to deviate from them now.

2 Historical reconstruction: an initial attempt within the Outer Eastern Islands region of the Solomon Islands

One set of investigations I am conducting is a holistic enquiry focused on that part of the south-east Solomon Islands that lies within Remote Oceania. Geopolitically this region now belongs within the Solomon Islands nation-state. Yet, in culture historical terms this group of islands and atolls should be set apart from the rest of the Solomon Islands archipelago for historical reconstructions and treated as a separate entity — the Outer Eastern Islands (OEI) of the Solomons. These islands and atolls also need to be discriminated from the rest of the south-east Solomon Islands that encompass the most eastern of the myriad islands and archipelagos all of which are properly assigned to the vast region of Near Oceania.

This OEI region is one Andrew and I have discussed many times, most recently in a long series of email exchanges and telephone conversations, none of them identifying a satisfactory resolution. The puzzle has been that presented by the linguistic situation in the OEI today, as mapped in Figure 1. It is clear enough that one cluster of current languages in that region are Outlier Polynesian and derive from one or more subgroups of the Nuclear Polynesian languages situated within the region of Western Polynesia. It is also uncontested that the speaker populations represent fairly recent arrivals within the region. A second cluster involves the three distinctive languages for each of the small high islands of Utupua and Vanikoro. A few researchers, such as François (2006), doubt that they are truly Oceanic Austronesian (Pawley 2006:241), and until recently most Pacific linguists
have placed them among the late stage Proto Oceanic (Poc) language subgroups of Remote Oceania (Lynch, Ross and Crowley 2002:208). But whether they might constitute a single subgroup or two separate ones, or even how they may relate historically to other proposed subgroups under POc, is not particularly well understood. Finally, there are the languages of the large island of Nendō or Santa Cruz, and those related languages spoken today on the Main Reef Islands, whose status as Oceanic Austronesian with multiple later borrowings, or non-Austronesian with a heavy dose of Polynesian and other Oceanic borrowings, has until now remained in dispute (Pawley 2006:218–219).

One part of the overall OEI region puzzle is the linguistic situation of the languages encountered on the islands of Utupua and Vanikoro. In their overview of the classification of Oceanic languages, Lynch et al. (2002:112) concisely summarise the various strands of the problem, without resolving them:

The islands of Utupua and Vanikoro each have three Oceanic [Austronesian] languages ... Perhaps more significantly, however, the six languages show an unexpected measure of diversity for the size and proximity of the islands, and, although we can recognize an Utupua family and a Vanikoro family, there are seemingly no innovations which allow us to attribute all six languages to a single group, let alone relate them to the Southeast Solomonic family, or to the Southern Oceanic linkage.

The differences between these languages and neighbouring groups indicate that they constitute one or two first-order CEOc [Central Eastern Oceanic] subgroups, whilst their linguistic diversity suggests that they have been in their rather isolated present location for a long time. It is possible that they have been there since the original dispersal of CEOc ...

This essay now endeavours to evaluate the several strands of the Lynch et al. overview above, by providing evidence in support of some of their observations, explanations for others, and alternatives to yet other strands along lines that they chose not to explore.

3 Biogeographical considerations

The hypothesis that the number of languages spoken on any given island in Melanesia is to a high degree correlated with island size, as well as with environmental diversity, was advanced in the mid 1970s (Lensink 1976; Terrell 1974). However, Pawley and Green (1984) saw much of the language or dialect diversity within the Austronesian languages of Island Melanesia as being quite recent, and — particularly in Fiji and Polynesia — also dependent on factors other than island size alone (Pawley 1981).

Pawley (2007) has now delved further into the question ‘Why do Polynesian island groups have one language and the Melanesian islands have many?’ His discussion begins with an outline set of explanations developed by Bellwood (1995:101–103) for the sudden, very expansive Austronesian diaspora that largely took place between 4000 and 2500 years ago. For this phenomenon, seven factors are singled out as underpinning the rapid dispersals of speakers of these languages, because each conferred clear advantages over the indigenous populations of lands they settled. The advantages divide into those of a cultural (i.e. technical and/or economic) sort, or those that draw on certain cultural attributes (i.e. values and ambitions) that served as much stronger motives for exploration and colonisation. These and other related factors are then examined in greater detail for the region of Remote Oceania. Pawley concludes that when Lapita voyagers into this region
found a new island group ... [these] colonists rapidly explored and lightly settled it. This initial colonising phase was typically followed by a sequence of demographic, economic, and technological developments.

Population growth led to a denser distribution of settlements, more intensive agriculture, more local marriage and trade, weakening of kin ties with distant communities, and linguistic divergence. This sequence of events happened in both Southern [Island] Melanesia and West Polynesia and Fiji, but it happened more slowly in West Polynesia/Fiji.

In the latter region, the maintenance of long distance voyaging, both within island groups and between neighbouring island groups, in the view of Pawley (2007) should ‘be attributed in large part to the rise of powerful chiefs.’

In contrast to this strongly cultural viewpoint as an explanation, Lensink explored the issue for the area of Island Melanesia (thus excluding Fiji and Polynesia) from the perspective of correlations with six environmental factors: (1) maximum elevation, (2) limitations on the availability of marine food resources measured by three quantifiable parameters, (3) mean annual rainfall, (4) mean annual temperature at sea level, (5) fraction of land area in Andosols, and (6) fraction of area in Rendzinas soils. Three of these environmental variables proved not statistically significant in predicting the mean size of any given land area occupied by a single language. The three that did account for 83 percent of the variation within a calculated ‘mean language land area’ were then subjected to closer examination. The one related to the imposition of restrictions on the availability of marine resources to any given land area proved to have the most important direct influence on mean language land area. Two measurable indicators of terrestrial food sources also proved significant in predicting the size of the language area of any given group of speakers of a single language — the relevant amount of land in nutrient-rich ashfall soils or Andosols, and the relative amount of land in Rendzinas soils, most often encountered on raised coralline landscapes. A host of other variables are also noted that might be expected to influence the mean language land area and account for the remaining variance in individual cases. Finally, Lensink (1976:106), while allowing that cultural or non-environmental variables might also be just as important, did not develop this possibility; this is in rather marked contrast to later work on the topic by Pawley, or Bellwood. Instead, Lensink’s focus is the influence of terrestrial and marine food productivity sustaining the geographical extent of regional populations of horticulturists that are recognised as single ethnolinguistic units. In short, this research has ‘emphasised gross environmental differences and ignored cultural and demographic factors’ (Lensink 1976:115), making it less robust in mathematical terms and insufficiently balanced in its coverage of all the explanatory possibilities that may have come into play.

As a consequence, other considerations that relate to historical, social and additional biogeographical factors should be acknowledged and explored in depth. Terrell (1986:55), for example, sets out the relationship in the Solomon Islands between land area in square miles and the number of languages spoken on fifteen islands: this is based on a study he published in 1977. A year earlier, Lensink (1976:Table 1) had documented much the same kind of relationship between land area and number of languages spoken on each island by means of a survey of 90 island units; these units extended from islands off the north coast of the Papua New Guinea mainland, through the islands of northern Island Melanesia, to those of southern Island Melanesia ending with New Caledonia and the Ile des Pins. His survey parameters excluded the vast and varied, large and small islands cluster of Fiji, even though most scholars at that time would have placed the Fiji group within the geographic
region assigned to Island Melanesia. The easterly end point of Lensink’s survey could be viewed as a pragmatic decision on his part, since the linguistic situation in Fiji was incompatible with the Island Melanesian model he had developed, and required other explanations along lines that have been argued by Pawley decades ago and yet again more recently (1981, 2007).

Moreover, it is also the case that other factors affecting rates and diversification in the various language subgroups (including the different kinds of diversity among languages — genetic or historical versus language and dialectical divisions — social practices contributing to maintaining identity, and economic variability reflecting varied environmental considerations) are all potential contributors to the equation (Terrell 1986: 246–247; Pawley 1981). Therefore, researchers must always be mindful of factors other than sheer island size, and several of these rapidly emerge when reviewing the OEI region of the Solomons.

With this as background, consider the overall findings of Lensink in respect to this region. Tikopia, Anuta and the Duff Island group (which includes the still inhabited island of Taumako) are not listed in his table although, as research on his part would have predicted, they reveal only one language spoken on each. In discussing the issues in Polynesian Outlier prehistory, Irwin makes a pertinent biogeographical observation, that when these entities are subjected to systematic comparative analysis with the larger nearby island groups of Melanesia, with few exceptions, the Outliers are all geographic isolates among their non-Polynesian neighbours, which provides the main common factor other than the linguistic and cultural affinities that link them together. Their systematic relationship is actually with Melanesia and Micronesia, not Polynesia, and confirms an earlier assertion that there is almost a case for renaming the Outliers. The survival of Polynesian traits, notably language (emphasis added), occurred because they outlie from Melanesia and Micronesia, while their spatial relationship to Polynesia is immaterial. (Irwin 1992:187)

When the relationships Irwin (1992:Figure 71) is examining are presented in graph form, the caption notes: ‘Many Melanesian islands may have had complex histories of Polynesian and other outside influence, but the evidence for it survives best with isolation’ (Irwin 1992:186).

In the OEI, Lensink does not enquire into either the situation on the atolls of the Outer Reef Islands of Nupani, Nukapu, Pileni and Matema, or that of the raised coral island of Niniloli, all of which are now occupied by upwards of a thousand speakers of the Vaeakau dialect of Vaeakau-Taumako (formerly known as Pileni). A similar number of speakers of Taumako, another dialect of this Outlier language, reside in the high islands of the Duff group (Hovdhaugen 2006).

However, the tiny atoll islands along the Outer Reef lagoon have probably only emerged and become habitable in the fairly recent past (Dickinson 2001, 2003). This is in keeping with the observations of other geologists and geoarchaeologists that the ashfall-derived topsoils of the region were restricted to the main raised coral islands lying to the east of the Outer Reef Islands. Hence these ashfall soils were not observed covering landscapes on the atolls of the Outer Reef. Thus, when surveying to locate ancient sites, I recovered from their surfaces only the typical *Terebra*, *Mitra* and other shell adze types 9 and 11, of a kind that appear throughout this region after circa 1200 AD (Kirch and Yen 1982:221–232 and Table 29). The *Terebra* shell adze in particular is known to be a widely occurring artefact type useful as a chronological marker in Island Melanesian sites of the
last 700 years (Spriggs 1997:195, 205). On the temporal scale this type seems to appear slightly earlier in Eastern Micronesia, suggesting to a number of researchers that this kind of shell adze may have had its origin in that part of the Pacific.

In the extensive list on Lensink’s Table 1.1, the focus for the OEI is on the main Santa Cruz Group — Nendō, Lord Howe (Tömötö Noi) and Trevanion (Tömötü Neo) — for which three languages are listed as occupying a total volcanic and raised coral land surface of 212 km$^2$. The language distribution is discussed, mapped and supported by Green (1976b:51–52), although most investigators would reduce the actual number of languages among the multiple named dialects to just two — Natügu and Nagu [or Nanggu]). They form one part of the Reef/Santa Cruz subgroup, the other being the Reef Island language of Æiwoo. Lensink neglects the Æiwoo language, distantly related to the two on Nendō, which together constitute the Reef/Santa Cruz subgroup. Æiwoo speakers presently reside on the Main Reef Islands of upraised coral limestone that are now very fertile because during the last millenium they were covered in volcanic ash. Thus, on a smaller island area they are able to sustain themselves in larger numbers than can the inhabitants of Nendō who occupy a larger volcanic high island (Green 1976c:247–249). Furthermore, no mention is made of the Vaeakau dialect of an Outlier Polynesian language distributed across the scattered atolls or the one Main Reef small raised islet discussed above. Despite these obvious gaps, only two results within Lensink’s data set for the Outer Eastern Island region really lie beyond the overall size and environmental parameters he has found to obtain throughout Island Melanesia. In short, when the gaps are filled, most outcomes within that region prove entirely consistent with the generally expected pattern that his data predict.

The two outstanding exceptions are the islands of Utupua and Vanikoro. Both are well under 100 square miles (i.e. 259 km$^2$) in surface area, and therefore the kind of islands on which a single language is deemed the usual outcome. According to Lensink (1976:Table 1.1) Utupua’s surface area may be estimated as about 29 square miles (75 km$^2$) and that of Vanikoro as 65 square miles (168 km$^2$). However, the literature sources identifying the number of languages found on these two islands, and used in constructing the entries in Table 1.1 (Lensink 1976:32), prove misleading. In listing only one language present on each island, the outcome appeared quite in keeping with his model. Yet the reality is that each island has long contained three distinct languages. The history of language identification, from the time of described European contact in 1834 up to 1975, had been carefully reviewed by Green (1976b:55–59) at approximately the same time Lensink was independently carrying out his data collection.

Over the last three decades, following this review of the unusual linguistic diversity on Utupua and Vanikoro, I have been impressed by Lensink’s observations on the general effect of island size on the number of languages to be expected, and my knowledge of the certainty of three languages on each of these islands. It seemed that this situation clearly required explanations beyond those solely related to the size of the island. One explanation I favour is that the present-day situation reflects an outcome from closely related languages that were far more widely distributed on other islands within the OEI region in the past. Some speakers of these languages returned to islands where at least one of their linguistically and socioculturally related populations had always maintained a foothold. They came ‘home’, so to speak, having been displaced either (a) from the Main Reef Islands, or from some part of the islands which make up the Nendō Group where they have been replaced by Reef/Santa Cruz-speaking populations, or in other instances (b) from
various of the outlier islands more recently inhabited by Polynesian speakers, Vaeakau-Taumako, Tikopia and Anuta.

Evidence of (b) associated with a return to Vanikoro is the presence of loans from an Outlier Polynesian language in the three languages of that island. Andrew Pawley compiled one list each for Utupua and Vanikoro, consisting of 60 items that are highly stable in Oceanic (Pawley 2009) and I compiled corresponding lists for each of the other four languages, with the results set out here.

<table>
<thead>
<tr>
<th>Utupua Island Group</th>
<th>No. of retentions, POc lexical items</th>
<th>No. of borrowings from Outlier Pn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asu(m)bao</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>Nembao</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>Tanimbili</td>
<td>24</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vanikoro Island Group</th>
<th>No. of retentions, POc lexical items</th>
<th>No. of borrowings from Outlier Pn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vano</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>Tanema</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Buma</td>
<td>22</td>
<td>3</td>
</tr>
</tbody>
</table>

The items in Vano are stone (*faka*), sky (*teliko*) and dig (*fi*); in Tanema they are stone (*faka*) and sky (*telika*), and in Buma stone (*faka*), day (*temoro*) and sky (*telika*). In the longer lists I have assembled, many more such Outlier Polynesian borrowings from Tikopia or Anuta appear: the sources of these would be language speakers who undertook the back migration to Vanikoro.

If one starts with the conjecture that the recently discovered Oceanic subgroup of Temotu, comprising the non-Polynesian languages of the OEI, is one of the primary subgroups of the Oceanic branch of the Austronesian language family (Ross and Næss 2007, Ross et al. 2008) and that the dispersal centre of Oceanic was the Bismarck Archipelago (Pawley 2003, 2008; Ross 1988) it is reasonable to assume from the dated archaeological evidence that the initial speakers of Temotu entered the OEI region of the Solomons 3000 to 3150 years ago from a homeland much farther west (Green and Yen 2009:160–166). This implies that they voyaged more or less directly from somewhere in the Bismarck Archipelago and within that region the source communities of people speaking these languages can most likely be located within a zone comprising central and northern New Britain and southern New Ireland.

Using language phylogenies to reveal the expansion of Austronesian languages into the Pacific, Gray et al. (2009) have shown these can be modelled by a series of expansion pulses and pauses. Their Pulse 1 includes major language subgroups that begin in Island Southeast Asia and extend out into eastern Indonesia. From this latter region, Oceanic and — after its breakup — the resultant major subgroups derive. The Southeast Solomonic language groups, located within the Southeast Solomons of Near Oceania, is one of these subgroups, and in the adjacent OEI region to the east (lying within Remote Oceania) Temotu is the other subgroup of concern here. The languages which comprised Temotu subgroup in the Gray et al. (2009) definition have more recently split into two branches within their schema, nominated as the Reef/Santa Cruz and Eastern Outer Islands. The Reef/SantaCruz Islands include languages of Malo, Lwepe, Nanggu, and Aiwo (Gray et al. 2009:Figure S5). The Eastern Outer Islands include the languages of Teanu, Buma,
Tanema, Vano, Tanimbili, Assumboa, and Nembao. Overall their analysis shows that within Near Oceania, the South-east Solomonic subgroup (Pawley 1972) has strong support with a probability of 1.00. The Temotu subgroup is also strongly supported with a probability of 1.00 (Gray et al. 2009 SOM Text, p.5–6). In Figure 1 Gray et al. (2009) identify these Temotu languages as having an independently indicated age of circa 3500 years BP, corresponding well to their inferred age from archaeology and linguistic subgrouping.

Other biogeographical observations also need to be highlighted when assessing those factors crucial to understanding the culture-historical puzzle presented by the populations that became principal inhabitants of the OEI of the Solomons. The first is that the islands of this OEI region were the initial land masses from which southern Island Melanesia, eastern Micronesia, and Western Polynesia were colonised. The significance of a biogeographical boundary between Near and Remote Oceania to the human history of Pacific colonisation was explored over 34 years ago by Green (1991:491), without providing names for the distinction. A collaboration with Andrew Pawley shortly thereafter, and his assertion that conceptual analytical entities lacking names as shorthand identifiers were unlikely to ever be widely adopted, prompted the adoption of Near Oceania and Remote Oceania as distinguishing terms (Pawley and Green 1973; Pawley 2003:17–20, 2005:xvii). Green (1976a:11–16 and Figure 2) further explored the importance of this boundary, drawing on evidence from the recent past, and Spriggs (2000) more recently has produced insights on the subject in an essay on the Solomon Islands as bridge and barrier in the settlement of the Pacific. One point often neglected is that from the time of the first colonisation of Remote Oceania, and over the three millennia since, there have been human contacts across this long-standing ecological boundary. In terms of Oceanic Austronesian voyaging and the opportunity to invent additional navigational techniques in crossing it, the separation is perfectly understandable (Irwin 1998:198–199 and Figure 2). As Irwin writes, ‘... there was ocean to be crossed, not only with colonisation but in all later interisland contacts in any direction’.... [thus] there is a need to invoke ‘seascapes’, and the smallest of these in Remote Oceania was that of Santa Cruz or the OEI region.

The first ocean voyages out from Near Oceania to the Santa Cruz Islands in Remote Oceania were indeed two-way, and archaeological evidence shows that obsidian from the Bismarcks and chert from Ulawa in the south of the main Solomons chain continued to arrive in Lapita settlements in the Santa Cruz group. ... [The maps for this part of the Pacific make clear] that Nendö and the Reef Islands lie due east of Makira, and Taumako is east of Ulawa. It is interesting that the same zenith stars passed over the south of the main Solomons chain as over the Santa Cruz Islands and this can hardly have escaped the attention of the people who lived there. To travel between the south of the main Solomons and the Santa Cruz group was as simple as following a zenith star path — east or west — with the seasonal winds. Thus, from the very beginning of voyaging in Remote Oceania there was a simple and obvious opportunity to begin to be aware of the navigational technique of latitude sailing. (Irwin 2006:73)

The implications of the above set of observations are all too evident. The Southeast Solomon Islands and their well marked and distinctive subgroup of conservative Oceanic Austronesian languages will have always played some kind of role in the cultural history of the OEI.
Irwin’s other biogeographical observation, again based in the voyaging capability of Remote Oceanic Austronesian speakers, also bears repetition: ‘Santa Cruz was the most successful [computer simulated] origin for canoes that reached [eastern] Micronesia …’ (Irwin 1992:147), ‘while the origins of eastern Micronesia [based on other simulations] could have involved multiple settlement from eastern Melanesia and possibly Fiji/Western Polynesia’ (Irwin 1992:132, with simulation results discussed in detail on pages 120–123).

Therefore, when the atolls of eastern Micronesia were settled shortly after the beginning of the first millennium AD, once they first became available for human habitation (Dickinson 2001, 2003), the archaeological evidence provides dates for their colonisation (Weisler 1999, 2000, 2001). Thus the region of eastern Micronesia has also played a role in the cultural history of the OEI of the Solomons, and it is therefore necessary to take contacts with that part of Micronesia into account.

Finally, under the rubric of biogeography, other internal environmentally driven changes in the region require consideration. The highly probable emergence from the sea, during the last millennium or so, of the tiny atolls of the Outer Reef Islands now inhabited by Vaeakau-Taumako speakers of Outlier Polynesian has been mentioned above. At present, although recent geological events involving uplift have not been detected on the various sized islands among the Main Reef Islands group (Craig 1981:41, Figure 3; Green et al. 2008) — which might have affected their habitation history — such uplift had certainly occurred in the millennia before they were settled. More important to their human history of occupation is that some 1000 years or more after they were settled, a major volcanic tephra event, sourced to the still active volcanic island of Tinakula, blanketed these islands in 20 to 30 centimetres of ash fall (Wall and Hansell 1976). Although a more precise time interval for this event remains undetermined through direct radiometric dating, all lines of evidence point to a short interval at some point in the first millennium AD.

The same volcanic ash fall event also covered the western portion of the high island of Nendö, extending eastward as a diminishing dusting no longer detectable as a discrete layer (Green et al. 2008). On Nendö, however, changes to the landscape in the last 3000 years through repeated uplift events among its many fault blocks, and occasional submergence to below modern sea level in at least two places, also impacted the history of its inhabitants. This too becomes important in interpretations of remains recovered to date by archaeology, and representative of the last 3150 years.

Each of the volcanic islands now occupied by Outlier Polynesians has proven to have a long and complex history of up to 3000 years (Kirch 1982; Kirch and Yen 1982; Leach and Davidson 2008), just as Irwin (1992:187) anticipated. Hurricane-induced tidal surges and vegetation devastation are known to have had an effect on both sites at Kahula (on Taumako), and the successive deposits underlying the location of the current main village on the island of Anuta. Similar effects due to similar kinds of natural events are also known for Tikopia and its central sheltered lagoon, which in time became a freshwater lake confined by a (now inhabited) tombolo sand bar across the breached side of this old volcanic landform. Moreover, the investigation of its 3000 year history by archaeological survey and targeted transect excavation thoroughly demonstrated a whole succession of changes in the shape and area of its coastal landscape and the later formation of the current freshwater lake called Te Roto (Kirch and Yen 1982:325–334 and Figure 124).

In contrast to all these internal perturbations documented as having occurred on one or more of the above islands over the last 3150 years, geological and natural histories for the islands of Utupua and Vanikoro during this period have so far proven singularly
uninformative. Only the general aspects of their physical geology as extinct, partially submerged volcanoes are known. Utupua is an extinct, subsiding Pliocene volcanic cone with one older excentric crater and a younger, larger centrally situated crater. Vanikoro is an extinct multiple volcanic complex of Plio-Pleistocene age consisting of four main volcanic centres (Utupua Sheets EOI5 and EOI6, 1976 and 1977). There is, however, not enough information in the geological reports on these two islands to meaningfully correlate with their rather meagre to almost nil archaeological record of the last 3200 years. Certainly, during this period they did not escape the devastations of periodic tropical cyclones, coastal landscape accretion following their settlement, or the impact of earthquakes. However, precisely how any recent changes to the 75 km$^2$ and 168 km$^2$ land areas of Utupua and Vanikoro may have been a factor in each island now exhibiting three well-differentiated languages, rather than the expected one, is moot. Those considerations require additional multidisciplinary fieldwork of an intensive kind not currently viewed as urgent. This is unfortunate, for a more certain solution to the current puzzle of language associations with archaeological sequences and biological attributions in this region will almost certainly depend upon at least one project with a focus on those two islands in respect to each of these subdisciplines.

4 Conclusion

Of necessity, resolving the puzzle presented by the OEI region has to begin with some kind of conjecture of just how the historical linguistic comparative data, dated archaeological evidence and the outcomes of biological analyses are intricately related. A choice was made to focus in the first instance on the ‘aberrant’ linguistic situation on the islands of Utupua and Vanikoro (Pawley 2006). This proved to yield a set of postulates that seems to have theoretical merit. Certainly it goes some way toward solving one aspect of that puzzle already touched on by linguists like Andrew Pawley. Its basic postulates are these:

4.1 All islands of the OEI region were colonised by diuternal settlement populations (Green and Green 2007:253) between 3150 years ago and circa 3000 years ago.

4.2 These populations all spoke a language now recognised as one of the primary subgroups within the Oceanic subgroup assigned to the widespread Austronesian language family. The Temotu subgroup is now so strongly attested by multiple lines of evidence that among Pacific historical linguists this linguistic entity is not going to be easily overturned by contrary claims.

4.3 Extensive evidence that I continue to assemble points to Vanikoro as the source of the diuternal settlement populations of Tikopia and Anuta who colonised these small and somewhat remote high islands not long after their parental community was well established in Vanikoro. This is the most obvious source population befitting their geographic position and is consistent with the ethnographic evidence of the exchanges with Tikopia in particular and Vanikoro over many centuries (Firth 1961:136–143, 158–161). The archaeological evidence of the zebraic chalcedony from the Nanggu Lapita site of SE-SZ-8 on Nendō recovered in the early Lapita assemblages of Tikopia adds further strength to this deduction (Sheppard 1996:108).
4.4 Also, as befits the nearer geographic position of Utupua to Nendö, then Main Reef Islands, it would appear that this sinking set of islets was also colonised by biutermal settlement populations at about the same time as those in the Reef/Santa Cruz group, that is 3150 to 3000 years ago. The significant difficulty is that the earliest sites may well now be substantially under water; the earliest Lapita sites identified on Vanikoro so far have proven to be partially under water, and at times such sites are almost impossible to find. Through migration from Utupua, probably via the islands of the Reefs then above high water, Taumako was colonised with its own biutermal settlement population. Here, proof is furnished by the archaeological evidence from the 3000 year old Lapita site of Nenubo, situated within the Main Reef Islands, with potsherds that derive from those manufactured in Taumako in addition to coralline chert found in the Nenubo site that also derives from Taumako (Leach and Davidson 2008:295–296; Sheppard 1996:108).

4.5 After circa 1200 AD when the Temotu speaking populations of Taumako, Anuta and Tikopia had diverged from Proto Temotu to become wholly separate languages, Outlier Polynesian immigrants from the east arrived in quick succession to settle on these long inhabited small islands. These new immigrant groups came to dominate those islands linguistically within a short period of time. They also contributed new cultural elements that have been recovered archaeologically, ones with ultimate origins in West Polynesia. However, biologically their populations remained a composite, derived in large part from the prior inhabitants with the addition of new genetic material traceable to the Polynesian incomers from the east.

4.6 It was in this period that small populations from these Outlier communities, now including Vaeakau from the Outer Reef Islands region, made back migrations to their former home islands. Extensive borrowings of lexical items from the Outlier Polynesian languages (not presented here) suggest which two of the three Utupua languages best mark this back migration scenario and which of the three was the one that never moved. The extent of these borrowings is only weakly attested in the data assembled to date, yet, it would seem that there were back migrations to Utupua from both the Vaeakau zone as well as from Taumako itself. I believe there is probably sufficient evidence within my files to give this proposition a reasonable degree of cogency, and with additional effort more such evidence could be assembled. In the same manner, two of the three languages on Vanikoro, with extensive borrowings from Tikopia and Anuta in their lexicon, are marked as small populations who engaged in back migration to the Vanikoro cluster of islands. They differ significantly in the extent of borrowing from the one language that endured on that island group from the beginning of this island group’s settlement.

4.7 Finally, the biological mtDNA evidence from studies conducted by Friedlaender et al. (2002) and Friedlaender et al. (2007:66–67), drawing on laboratory-stored blood samples from the Reef/Santa Cruz Islands populations, reveal these results for the 69 samples that have been fully analysed. Twenty-three of the outcomes are B4a1, five are P1 and one is P4, twelve are Q1, seven are Q2 and ten are M27 (five are M27a and five M27c). There is also one M28a3. These outcomes are significant in respect to the composition of founding populations on the OEI. Thus, some 33 percent is composed of the Polynesian motif (B4a1) and could have had two principal sources (a) from the initial founding population and (b) from the more recent immigrants from Polynesia. At present these two most likely
sources of this Polynesian motif cannot be distinguished from each other. However, those
that are P1 and P4, Q1 and Q2, and M27, all strongly suggest these initial immigrants to
the OEI region from the west had already been sufficient time in the central northern New
Britain/southern New Ireland home area for them to have intermixed there with the prior
non-Oceanic inhabitants within the area who had preceded the Oceanic speaking Lapita
immigrants in that region. Only after those events had taken place did these mixed genetic
populations spread to the Reef/Santa Cruz region and thence into Southern Island
Melanesia.

Further exploration of such a proposition would require detailed investigations and new
samples from throughout the OEI region to establish it as the most likely explanation of
their current genetic makeup.

References

Biggs, Bruce, 1972, Implications of Linguistic Subgrouping. In R.C. Green and Marion

Bellwood, Peter, 1995, Austronesian Prehistory in Southeast Asia: homeland, expansion
and transformation. In Peter Bellwood, James J. Fox and Darrell Tryon, eds
The Austronesians: historical and comparative perspectives, 96–111. Canberra:
Department of Anthropology, Research School of Pacific and Asian Studies, The
Australian National University.

Craig, P.M., 1981, Nendö Island. In G.W. Hughes, P.M. Craig and R.A. Dennis, eds
Islands Government, Ministry of Natural Resources.

Davidson, J.M., 1999, Roger Curtis Green. In Tim Murray, ed. Encyclopaedia of

Dickinson, W.R., 2001, Paleoshoreline record of relative Holocene sea levels on Pacific

—— 2003, Impact of Mid-Holocene hydro-isostatic highstand in regional sealevel on

Firth, Raymond, 1961, History and traditions of Tikopia. Wellington: The Polynesian
Society.

François, Alexandre, 2006, Are Vanikoro languages really Austronesian? Paper read at

Friedlaender, J.S., F. Gentz, K. Green and D.A. Merriwether, 2002, A cautionary tale
on ancient migration detection: mitochondrial DNA variation in Santa Cruz Islands,

Friedlaender, J.S., F.R. Friedlaender, J.A. Hodgson, S. McGarth, M. Stoltz, G. Koki,
T.G. Schurr and D.A. Merriwether, 2007, Mitochondrial DNA Variation in Northern
Island Melanesia. In J.S. Friedlaender, ed. Genes, language and culture history in the

Gray, R.D., A.J. Drummond and S.J. Greenhill, 2009, Language, Phylogenies Reveal
text at www.sciencemag.org/cgi/content/full/323/5913/479/DC1.


Hollyman, Jim and Andrew Pawley, 1981, Bruce Biggs: a foreword. In Jim Hollyman
and Andrew Pawley, eds Studies in Pacific languages and cultures in honour of Bruce

Hovdhaugen, Even, 2006, A short dictionary of Vaeakau-Taumoko language. Oslo:
The Kon-Tiki Museum Occasional Paper 11.

Cambridge: Cambridge University Press.

—— 1998, The colonisation of the Pacific Plate: chronological, navigational and social

—— 2006, Voyaging and Settlement. In K.R. Howe, ed. Vaka Moana: voyages of the

Kirch, Patrick V., 1982, A Revision of the Anuta Sequence. Journal of the Polynesian

—— 2000, On the Road of the Winds: an archaeological history of the Pacific Islands
before European contact. Berkeley: University of California Press.

Kirch, Patrick V. and D.E. Yen, 1982, Tikopia: the prehistory and ecology of a Polynesian
Museum.

Journal of Archaeology Special Publication.

Lensink, Stephen C., 1976, Environmental factors affecting the areal size of regional
populations in Melanesia. Unpublished thesis, Master of Arts in Anthropology Degree,
University of Iowa.

Lynch, John, Malcolm Ross and Terry Crowley, 2002, The Oceanic languages. Richmond,
United Kingdom: Curzon Press.

Pawley, Andrew, 1966, Polynesian languages: a subgrouping based on shared innovations

—— 1972, On the internal relationships of eastern Oceanic languages. In Roger C. Green
and Marion Kelly, eds Studies in Oceanic culture history, vol. 3, 1–42. Honolulu:
Bernice Pauahi Bishop Museum.

—— 1981, Melanesian diversity and Polynesian homogeneity: a unified explanation for
language. In J. Hollyman and A.K. Pawley, eds Studies in Pacific languages and
cultures in honour of Bruce Biggs, 259–310. Auckland: Linguistic Society of New
Zealand.

—— 1991, Notes on a large small game hunter. In A. Pawley, ed. Man and a half: essays
in Pacific Anthropology and Ethnobiology in honour of Ralph Bulmer. Auckland: The


and C. Renfrew, eds Examining the farming/language dispersals hypothesis, 251–273.
Cambridge: McDonald Institute of Archaeological Research, Cambridge University.


18 Count not on substrata

PETER C. LINCOLN

Why is the alphabet in that order? Is it because of that song?
Stephen Wright

1 Introduction

This contribution seeks to clarify in five steps the use and derivation of counting words in Banoni. The first step reveals the multiple patterns of apparent prefixes. The second step reveals how the system logic reflects a particular method of counting on one’s fingers. The third step reveals the pattern of replacement for the word for 3 among nearby languages. The fourth step reveals a different pattern of replacement of the word for 10. The fifth step partially reveals the relation of Banoni paraphrasing to other counting systems used in nearby communities with a focus on the non-Austronesian South Bougainville family (SB) and the non-Austronesian North Bougainville family (NB). Even with gaps in the data, this last step establishes that any influence from those neighbouring languages cannot have been a simple process. Given that the systems of paraphrase are so diverse, it is simpler to conclude that Banoni paraphrasing for 6,7,8 results from creative forces within the language and not from the pressure of neighbouring languages. The notion that peculiarities in Banoni or closely related Piva are due to Papuanisation is undermined again, as it was in Lincoln (1976a). Also the notion that peculiarities in counting words in Remote Oceania are due to Papuanisation is weakened by implication.

1.1 Geographical context

As shown on the map below, Banoni and Piva are spoken on the west coast of Bougainville. These two form the Banoni-Piva family (BP) (Ross 1988; Lincoln 1976a, 1976b; Lynch et al. 2002). Together with the Oceanic languages from Nehan to New Georgia they form the Northwest Solomonic (NWS) linkage. In the general area there are several more languages. Some distance to the east are several Polynesian outlier languages

---

1 Special thanks to Bob Chandler, Joel Bradshaw, Will McClatchey, and an unnamed reviewer for generously providing suggestions to improve this paper; thanks to Bob Blust for early access to his unpublished work; and daume marono to John Soveke, who taught me how to count on my fingers, and William Tagis; tapeng to Theresa and Janos Suagotsu for confirming data. To all of them all credit and no blame.

2 In this regard, Blust (2005) noted some peculiarities that make Banoni unique among the languages he has worked on.
(Nukuria, Takuu, Nukumanu, and Luangiu) that appear to have been too far off shore to have had recognisable impact on Banoni or any other languages of the area. On the island of Bougainville there are eight languages that are not Austronesian. Little comparative work has been done on these languages. The preliminary classification in the survey by Allen and Hurd (1963) has been slightly revised by Ross (2005) who recognises the NB family: Rotokas, Eivo, Konua, and Keriaka spoken north and inland of the BP area; and the SB family: Nasiot, Nagovisi, Motuna (Siwai), and Buin spoken east of the southern BP area. Beyond Bougainville to the east there are four surviving non-Austronesian languages. Again, little comparative work has been done on these languages. Todd (1975) suggested these four were related. Ross (2005) recognises this grouping as the Central Solomons (CS) family: Bilua, Baniata, Lavukaleve, and Savosavo. The NB, SB, and CS groups plausibly all share ancestry with people who were on and around Buka for some 28,000 years (Wickler 2001), giving them a good long while for their languages to diverge. Wurm’s (1975) suggestion that all the non-Austronesian languages east of Umboi and of the main island of New Guinea would reveal sufficient cohesion to justify his East Papuan Phylum remains largely unsupported, though serious testing is just beginning (Ross 2005; Dunn et al. 2005).

Map: Languages in the Banoni area
The NWS languages closest to BP in the south are the Mono-Alu Torau (MAT) family: Mono-Alu spoken on small islands south of Bougainville, Torau spoken on the east coast north of Arawa, and Uruava formerly spoken in the village of Arawa.

Currently there is language contact on various interfaces. For example, the matrilocal Banoni in the south have for generations married spouses from Nagovisi and Motuna. Banoni men are likely to be quite fluent in neighbouring languages. In-marrying men may eventually learn Banoni. The man in the exemplar case of dual-lingualism (Lincoln 1979) has now some thirty years on learned Banoni, perhaps urged on by the special circumstances of the Crisis (1989–2000) when it would have been advantageous for all members of the village to speak a language mostly unknown to outsiders. Women are much less likely to learn another local language. Traditional patterns of men knowing several languages while women stayed home speaking almost exclusively their home language are now fading with so many having learned Tok Pisin, or even English.

1.2 Conventions

Shared words not directly inherited are referred to as ‘copied’ following Thurston (1987). Unadorned numerals like 3 are used as a shorthand for phrases like ‘the word for three’ or ‘three’ or even ‘3’. POc forms unless otherwise noted are from Lynch et al. (2002).

Table 1: Counting words of the BP family and POc

<table>
<thead>
<tr>
<th></th>
<th>Banoni</th>
<th>Piva</th>
<th>POc (Lynch et al. 2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kadaken</td>
<td>kadaken</td>
<td>*ta-sa, *sa-kai, *tai, *kai</td>
</tr>
<tr>
<td>2</td>
<td>too-m</td>
<td>to-nua</td>
<td>rua</td>
</tr>
<tr>
<td>3</td>
<td>da-pisa</td>
<td>to-pisa</td>
<td>tolu</td>
</tr>
<tr>
<td>4</td>
<td>to-vatsi</td>
<td>e-vatsi</td>
<td>*pati</td>
</tr>
<tr>
<td>5</td>
<td>ghi-nima</td>
<td>niima</td>
<td>*lima</td>
</tr>
<tr>
<td>6</td>
<td>bena</td>
<td>e-bena</td>
<td>*onom</td>
</tr>
<tr>
<td>7</td>
<td>bena to-m</td>
<td>bena to-nua</td>
<td>*pitu</td>
</tr>
<tr>
<td>8</td>
<td>bena ka-pisa</td>
<td>bena to-pisa</td>
<td>*walu</td>
</tr>
<tr>
<td>9</td>
<td>visa</td>
<td>sia</td>
<td>*siwa</td>
</tr>
<tr>
<td>10</td>
<td>manogha</td>
<td>manogha</td>
<td>*sa-[ŋa]-puluq</td>
</tr>
<tr>
<td>20</td>
<td>manogha toom</td>
<td>manogha tonua</td>
<td>*rua-[ŋa]-puluq</td>
</tr>
<tr>
<td>100</td>
<td>rausu</td>
<td>nasusu kadaken</td>
<td>*Ratus</td>
</tr>
<tr>
<td>1000</td>
<td>kokoreo</td>
<td>nasusu manogha</td>
<td></td>
</tr>
</tbody>
</table>

1.3 Counting words retained

As can be seen in Table 1, Proto Oceanic (POc) has a complete decimal system in the sense that there is a separate word for each number from 1 to 10 even though the word for ten is composite, consisting of *sa ‘one’ + ŋa LIGATURE + *puluq ‘ten’. The system for BP is not so complete. Part of each BP word for 2, 4, 5 and 9 plausibly derives directly from the POc forms; the rest are innovated. Sound changes in Banoni obscure developments. The development of *rua > m is clear when we see the steps: prefixing to- to *rua, then *r > n reaching the modern Piva tonua, then in Banoni the final *a suffered irregular loss,

---

3 Hyphens in the Banoni and Piva forms highlight the presumed stem.
and then the final syllable nu shortened to m. This last step is paralleled in *poñu > ghom ‘sea turtle’ and *paqoRu > ghoom ‘new’.4 The change of *siwa > visa involves metathesis, which also occurs in *tasik > taghisi ‘sea’, *taliña-na > tangin-na ‘ear-his’, *buaq > bughava ‘betel nut’, and in the mouths of village children my wife’s name Satoko often came out as Sakoto. These developments are outlined in Table 2.

<table>
<thead>
<tr>
<th>POc</th>
<th>Gloss</th>
<th>Piva</th>
<th>Banoni</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>*rua</td>
<td>2</td>
<td>nua</td>
<td>m</td>
<td>final nasal rounded</td>
</tr>
<tr>
<td>*pati</td>
<td>4</td>
<td>vatsi</td>
<td>vatsi</td>
<td></td>
</tr>
<tr>
<td>*lima</td>
<td>5</td>
<td>nima</td>
<td>nima</td>
<td></td>
</tr>
<tr>
<td>*siwa</td>
<td>9</td>
<td>sia</td>
<td>visa</td>
<td>metathesis</td>
</tr>
</tbody>
</table>

1.4 Counting words innovated

As presented in Table 1, the data in BP has several kinds of overlapping innovations. The number 1 is a special case. Counting words function as such because of their order, but 1 by itself is unordered. The multiple reconstructions of 1 in Table 1 reflect difficulties with meaning (Lynch et al. 2002:73). BP kadaken is quite unlike the various POc reconstructions. That no language in the area has a form more like kadaken than the Mono-Alu kaala (see Appendix, Table A1) suggests that the Banoni innovation is internally motivated and not the result of language contact.

Taking the other innovations step by step; the BP counting words 2,3,4,5 all have added initial syllables independent of the rest of the word, the presumed stem;5 BP 3 and 10 have stems copied from other NWS languages; and POc 6,7,8 have been replaced entirely by paraphrase.

The first ten counting words from all the languages from Nehan to Savosavo are presented in the Appendix: Table A1 has data from the Oceanic languages, Table A2 from the non-Austronesian languages. Actually data in the crucial 6,7,8,9 range are missing for Papapana in A1, and for Keriaka in A2. The rest of this paper draws on these data to present observations that will put the BP system into perspective both in terms of the present context is and what changes from prior contexts may be inferred.

2 Step one: initials

Throughout the family, many counting words have a syllable before the more familiar numeric stem. These apparent prefixes need to be identified to analyse the systems, but they do not seem to pattern in any useful way. For example, in Banoni, the apparent prefix to- seems to signal ‘numeric’ in toisa ‘how many’ contrasting with ‘temporal’ in noisa ‘when’ and both words appear to reflect earlier *pica(n) ‘how many’ (Lynch et al. 2002:72). However, not every BP counting word has the to- prefix. Indeed, the beginning syllables of Banoni counting words are chaotic. This chaos reflects the wider patterning in the NWS as a

---

4 Actually, in Banoni syllable final nasals are not distinguished; following the round vowels o and u the final nasal is pronounced m. Similarly, before the round vowels o and u the voiced bilabial fricative v and the voiced velar fricative gh are not distinguished, indeed both are said together.

5 The ka- of kadaken may also fit this pattern except that the stem remains obscure.
whole. There are some localised tendencies: to- in the north, ka- on Lauru and around New Georgia, and e- in between. Banoni shares to- and ka-, but not da- and ghi-. Piva shares to- with many, but shares e- only with Torau and Mono-Alu and not with Banoni. Only with 2 and 7 do Banoni and Piva have the same prefix with the same stem. In the compound bena tom 7, the prefix does not occur at the beginning of the compound suggesting that the whole is no longer analysed as separate parts. In the compound bena kapisa 8, Banoni uses a different prefix, while Piva bena topisa repeats the word for 3 unchanged.

A closer look at these chaotic prefixes only reveals more questions. The e- prefix in Piva e-vatsi 4 and e-bena 6 is apparently different from the article e marking singular non-human nouns, which contrasts with a that marks singular nouns as human. If the prefix is the article, why do only 4 and 6 need to be marked as singular nouns? Aren’t they actually plural? The da- prefix in da-pisa 3 is similar to the accreted initials on Banoni words that might otherwise begin with a (after the loss of initial *q): date < *qate ‘liver’; dzai < *api ‘fire’; daso < *qasu ‘smoke’; dzaru < *aRu ‘tree, Causuarina equisetifolia.’ However, other evidence for **apisa as the stem is lacking. The stem pisa occurs in 8 with ka- prefix. Of the more than sixty words in Banoni that begin with the syllable ka only kasi- ‘sibling of the same sex’ reflects a known POc source *tasi- but with an irregular *t > k change. On Lauru, Ghone and Varisi the same form kapisa is only for 3; there kazalu 8 derives from *walu. The Banoni accretion to *lima in ginima 5 is without parallel within the NWS area; even without known parallel in Banoni itself. Of the few Banoni words that start with /ghi/ only ghinima has a known source.

Blust (forthcoming, §5.1.4) recently looked at a large sample of Austronesian counting word initials including some NWS forms discussed here, and concluded that the nature of the semantic field lent itself to having what Matisoff (1995) treats as ‘onset runs’ based on similar findings in Tibeto-Burman languages. In other words, the apparently regular use of to in Nehan 2–8, and of ka in Varisi 1–9 are each suggestive, but at the same time conflicting, and no reconstruction is implied.

### Table 3: Apparent prefixes on counting words in Bougainville and Lauru languages

<table>
<thead>
<tr>
<th>Language</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nehan</td>
<td>ka-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td></td>
</tr>
<tr>
<td>Halia</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td></td>
</tr>
<tr>
<td>Solos</td>
<td>me-</td>
<td>hue-</td>
<td>hue-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td></td>
</tr>
<tr>
<td>Petats</td>
<td>hua-</td>
<td>hua-</td>
<td>ho-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td></td>
</tr>
<tr>
<td>Hahon</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td>tau-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinputz</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td></td>
</tr>
<tr>
<td>Torau</td>
<td>ka-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td></td>
</tr>
<tr>
<td>Mono-Alu</td>
<td>ka-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td>e-</td>
<td></td>
</tr>
<tr>
<td>Vaghua</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td></td>
</tr>
<tr>
<td>Ghone</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td></td>
</tr>
<tr>
<td>Varisi</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td>ka-</td>
<td></td>
</tr>
<tr>
<td>Banoni</td>
<td>ka-</td>
<td>to-</td>
<td>da-</td>
<td>to-</td>
<td>ghi-</td>
<td>(to-)</td>
<td>(ka-)</td>
<td>(to-)</td>
<td>(ka-)</td>
<td>7,8 prefix on 2,3</td>
</tr>
<tr>
<td>Piva</td>
<td>ka-</td>
<td>to-</td>
<td>to-</td>
<td>to-</td>
<td>(to-)</td>
<td>(to-)</td>
<td>(to-)</td>
<td>(to-)</td>
<td>(to-)</td>
<td>7,8 prefix on 2,3</td>
</tr>
</tbody>
</table>

---

6 Matisoff (1995) mentions examples in Indo-European. Indeed, consider English ‘four’ and ‘five’ also ‘six’ and ‘seven.’
3 Step two: internal logic

The BP languages do not have complete decimal systems in the sense that POc *onom 6, *pitu 7, and *walu 8 were replaced by paraphrase. The internal logic is that bena 6 in other contexts means ‘cross over (to the other side)’. That same meaning applies when Banoni count on their fingers. One starts with the left hand open, palm toward the face, then the extended right index finger is placed against the little finger of the left hand and the count is ‘kadaken’, then the right index finger is moved to the left ring finger ‘toom’, then to the left middle finger ‘dapisa’, then to the left index finger ‘tovatsi’, then to the left thumb ‘ghinima’. Then one changes hands. The extended left index finger is placed against the right little finger ‘bena’, then moved to the right ring finger ‘bena toom’, then to the right middle finger ‘bena kapisa’, then to the right index finger ‘visa’, finishing with the right thumb ‘manogha’. Without this knowledge, the counting seems to progress unusually and illogically as 6, 6+2, 6+3, 9, 10. (Blust 2005). The clue to the BP logic was independently suggested by Glendon Lean in his study of Papua New Guinea number systems: ‘6 should be “bena kadaken” unless “bena” indicates that counting has transferred to the first finger of the other hand’ (Lean 1991:59). Lean was familiar with the counting technique, but lacked relevant details of Banoni. Herewith both are in print for the first time.

Given the shared detail of counting systems in Piva and Banoni, and the close relation of the languages overall, it is likely that the innovation of using the verb ‘cross’ to replace only the words for 6,7,8 happened once in this subgroup. Whether this change happened only in this subgroup is discussed later. Without further study of both languages, it remains uncertain whether the innovation occurred before BP became separate languages or whether it occurred in one and was copied by the other.

4 Step three: the stem for 3

As can be deduced from data in Table A1, the BP stem pisa 3 is shared exclusively by only some — not all — of the NWS languages of Lauru and of Bougainville in a way that implies copying rather than shared innovation supporting a subgroup. The stem pisa 3 occurs in some languages to the north: Halia, Solos, Petats, and Sapos, but not in Nehan, Teop, Tinputz, Hahon, or Papapana; also in Torau and Mono-Alu, but not in Uruava; and in some Lauru varieties8 Ghone, Varisi, Ririo, Lōmaubi, and Avasō, but not in Vaghua, Babatana,9 or Sisia. There is minor variation throughout this range: the westernmost varieties lack the final vowel, hence pis; the easternmost show a different first vowel, pasa. No similar form for 3 has been found elsewhere in the Pacific.10 Indeed, the only recorded language that has exactly pisa without a prefix is Ririo. When and how the copying actually took place is beyond the scope of this investigation. The languages that use a variation on pisa copied it from others or inherited it from earlier copiers. The various forms on Lauru suggest that the word may have developed there and was later copied from Lauruans by people in Buka, BP, and MAT. This pisa isogloss supports the claim that the

---

7 For example, na ko ta bena koromo (1s COMPLETIVE NOW cross water) ‘I am crossing a river’.
8 Data from ten Lauru varieties are listed in Tryon and Hackman (1983) but only Vaghua, Varisi, Ririo and Sisia are listed as languages in Lynch et al. (2002).
9 A Babatana speaker was employed by Andrew K. Pawley in his first innings teaching linguistics.
10 In central Vanuatu, Nguna has pisa ‘few’ but for 3 POc *tolu is retained (Schütz 1969).
Torau, who pair with the Mono-Alu in using *pisa while the adjacent Uruava use a reflex of *tolu, recently (mid-19th century) moved from the Shortlands to south-east Bougainville (Terrell and Irwin 1972). The isogloss implies further that groups on Lauru had contact with groups on Bougainville.

5 Step four: the stem 10

The stem *manoγa 10 has a distribution somewhat different from that of *pisa, hinting at traditional contact over a slightly wider area. To the north, Nehan Saposa, Hahon, Tinputz, and Teop reflect POc *sa-[ŋa]-puluq. Halia, maloto, Solos manoto, and Petats malot might derive from *manoγa, but in the absence of an explanation for the t, the resemblance is merely suggestive. To derive Papapana numanoa requires only the plausible weakening of γ and a prefix like accretion. The MAT family all reflect POc *sa-[ŋa]-puluq. The Lauru varieties except Vaghua all show forms easily derived from *manoγa. In the New Georgia area Lungga, Simbo, Nduke match *manoγa exactly and Ghanongga with a minor change to manŋa. Ughele, Roviana, Kusaghe, and Hoava have changed the vowels slightly more to maneγ. This pattern connects BP with parts of Bougainville, with most of Lauru, and with parts of New Georgia. The variation suggests a proto form of *manoγa with differentiation on Lauru and New Georgia.

6 Step five: relation to other systems

Now we address the complex question: Was the BP system characterised by the paraphrase structure {‘cross over’ INCREMENT} as part of only 6,7,8 motivated by contact with another language or group? As background, we consider the nature of the system and what we are looking for. Next we narrow the search geographically. Then we consider the patterns of a dozen potential languages. We will conclude that evidence for external motivation is lacking.

6.1 Clarifying the use of paraphrase

Most languages have counting words from 1 to 10 and start to paraphrase at 11, which is typically some compound of 1 and 10. Some other languages start at 6, typically some compound of 1 and 5. In the following sections I will avoid referring to the former as a ‘base ten’ or ‘decimal’ system and the latter as a ‘base five’ or ‘quinary’ system because these mathematical terms have different implications. In the mathematical sense, a system starts over again at multiples of the base. A decimal system begins using two digits at 10, three digits at the base times the base 100 (10^2), four digits at the base times the base times the base 1000 (10^3) and so on. A base five system changes to two digit mode at 5, to three digits at what we call ‘twenty-five’ (5^2), and changes to four digits at what we call ‘one hundred and twenty-five’ (5^3). As far as I know, all of the counting systems from Bougainville out into Remote Oceania are — in this mathematical sense — base ten or decimal. Many Oceanic languages reflect POc *Ratus 100, but I know of none with special

---

11 This logic is present in the familiar expression ‘six figure salary’ to mean more than $100,000, (base 10 to fifth power).
words for \((5^2)\) or \((5^3)\). No POc form has been reconstructed for 1000 or higher as the forms from different areas show little agreement.\(^\text{12}\)

Having the word for 5 the same as for ‘hand’ and does not imply a ‘quinary’ system. Almost all Austronesian languages have something like POc *lima for both meanings and nearly all are decimal. Even if 6,7,8,9 are made up of compounds involving the word for 5, the system is likely to be decimal. If the word for ‘man, human’ is also used for 20, the system may still be decimal.\(^\text{13}\) So when Firchow (1987:46) declares that ‘the [Rotokas] counting system is quinary (as are the other Non-Austronesian languages on Bougainville)’, he is most likely just asserting that these languages use paraphrase in counting 6,7,8,9. In the Solomons chain from Nehan to Makira, the only languages that paraphrase for 6,7,8,9 are on Bougainville. All the languages from Mono-Alu to Makira even the surviving non-Austronesian CS languages have distinct counting words one to ten. Thus, the only modern languages that might be models for the BP pattern are those on Bougainville itself.

6.2 Subgrouping considerations

Subgrouping is used in this section merely to help organise the data. Nonetheless, it is interesting that very closely related language varieties may differ radically in their approach to paraphrasing. Buin and Motuna are closely related languages, but Buin uses a rare decremental paraphrase and while Motuna uses an incremental paraphrase. Even more radically, the Tinputz list collected by Lean uses incremental paraphrase and a list collected by Blackwood (1935:31) uses retained words in counting 6,7,8,9.\(^\text{14}\)

6.3 Getting at the syntax of paraphrase

The paraphrase techniques used to count between 5 and 10 on Bougainville are varied. The Oceanic NSL languages surrounding the non-Austronesian families NB and SB relate to each other as a loose linkage of ten languages and two smaller co-ordinate families of two and three languages respectively (Lynch et al. 2002:883–884). The details of the subgroups and their counting are summarised in Table 4.

\(^\text{12}\) In the Banoni area, as Lean (1991:65) noticed, there is a curious association of 1000 with ‘chicken’. Buin \textit{kukurei} and Motuna \textit{kukuraku} have both meanings (Laycock 2003:69; Oliver 1955:101). Banoni \textit{kokoreo} 1000 also means ‘chicken.’ Uruava \textit{kokolet}, Torau \textit{kokoleu}, and Nasioi \textit{kokore} 1000 (Lean 1991:65) have not been yet been confirmed with the meaning of ‘chicken’. In the Tryon and Hackman data, only Mono-Alu exhibits this coincidence \textit{ea kokolet} 1000 (1983:131) and \textit{kokolet} ‘chicken’ (1983:171).

\(^\text{13}\) The English word \textit{score} 20 does not affect the base of our system, nor does the word \textit{dozen} 12. The latter persists as an aggregate because it has so many divisors i.e. one can distribute a dozen donuts evenly among groups of 2, of 3, of 4, and of 6 people. Ten donuts serve well only groups of 2 or of 5.

\(^\text{14}\) Blackwood’s Kurtatchi data (1931/32; 1935) would seem to come from the same language as Lean’s Tinputz material, or at least the Hahon-Teop-Tinputz group. Both Müller (1954) and Allen and Hurd (1968) supply language maps indicating that the language of the Kurutatchi area is the Selau dialect of Halia, which links more closely to Buka languages than to Tinputz. However, the numeric data are as stated and indicate otherwise.
Table 4: How the Oceanic languages of Bougainville count:
paraphrase 6,7,8,9; use pisa 3; use mano‘a 10

<table>
<thead>
<tr>
<th></th>
<th>Paraphrase</th>
<th>pisa 3</th>
<th>mano‘a 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nehan-North Bougainville linkage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nehan</td>
<td>Nehan</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td>Solos</td>
<td>Solos</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td>Buka linkage</td>
<td>Halia</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td></td>
<td>Petats</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td>Saposa-Tinputz family</td>
<td>Taiof</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Saposa</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hahon</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinputz</td>
<td>both</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teop</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papapana</td>
<td>Papapana</td>
<td>&lt;no data&gt;</td>
<td>✓</td>
</tr>
<tr>
<td>Piva-Banoni family</td>
<td>Banoni</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Piva</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mono-Alu/Torau family</td>
<td>Mono-Alu</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Torau</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uruava</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.4 Details of paraphrase in northern Bougainville Oceanic languages

Of the languages in Table 4, the three northern Bougainville languages which will now be considered are presented in Table 5 with proposed analysis for paraphrase. As already mentioned (§6.2), the Tinputz data reported by Blackwood has number words for 6,7,8,9 that match well those of languages further north, but the numerals from 1–4 exhibit near identity with distinctive words only found in Tinputz, Hahon and Teop: compare Blackwood’s (1935) pe’es 1, boak 2, kukan 3, tana 4 with Lean’s (1991) paeh 1, bok 2, kukon 3, tana 4. Blackwood herself explains the variation as follows: ‘Numbers in most dialects go up to five. Where there are words for numbers from 6 to 10 they are similar in all dialects, and would appear to be borrowed’ (Blackwood 1931/32:209). In Table 5, Tinputz and Hahon exhibit a very similar pattern of paraphrase: \{5 LIGATURE INCREMENT\}. The 5 element and the increment element are both clear. In Tinputz mea functions as a ligature.\(^\text{15}\) Teop me/ mi ‘with, for’ (Mosel and Thiesen 2007) also has overlapping form and meaning similarities. All of these forms plausibly reflect POc *mai ‘with, and’. Hahon mana is slightly different in form but identical in use to the Tinputz form and thus plausibly another reflex.

\(^{15}\) Banoni has the similar form me-a with similar meaning ‘with-him’, i.e. comitative inflected for third person singular.
In Lean’s Teop data the LIGATURE appears to be \(ea\), but there are transcription errors. So, for example, Lean’s Teop 6 should read \((taonim\ me\ a)\ peha\ to\ toka\) and be analysed as ‘5 AND ARTICLE 1 RELATIVE perch’.\(^{16}\) Thus Teop too shows the reflex of POc *mai in the longer construction for 6,7,8,9, showing close agreement among all three of these languages. As Lean notes, the Teop forms have an extension (not reported for Tinputz and Hahon), meaning something like ‘on top’ or ‘more’, literally ‘which perches’. In terms of the pattern then, Teop extends Hahon and Tinputz pattern of \{5 LIGATURE INCREMENT\} to \{5 LIGATURE INCREMENT ‘more’\}, but many speakers actually abbreviate the long form to \{INCREMENT ‘more’\}. The Nasiö further south abbreviate in much the same way.

**Table 5:** Tinputz, Hahon and Teop\(^{17}\) paraphrase \{6,7,8,9\} with morpheme guesses

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tinputz</strong></td>
<td><strong>tonima mea paeh</strong></td>
<td><strong>tonima mea bok</strong></td>
</tr>
<tr>
<td></td>
<td>(5) LIG 1</td>
<td>(5) LIG 2</td>
</tr>
<tr>
<td><strong>Hahon</strong></td>
<td><strong>taunima mana paia</strong></td>
<td><strong>taunima mana buaku</strong></td>
</tr>
<tr>
<td></td>
<td>(5) LIG 1</td>
<td>(5) LIG 2</td>
</tr>
<tr>
<td><strong>Teop</strong></td>
<td>((taonim\ ea)\ peha\ to\ toka)</td>
<td>((taonim\ ea)\ buaku\ to\ toka)</td>
</tr>
<tr>
<td></td>
<td>(5) LIG 1 REL ‘perch’</td>
<td>(5) LIG 2 REL ‘perch’</td>
</tr>
<tr>
<td><strong>Teop(^{M&amp;T})</strong></td>
<td><strong>peha to-toka</strong></td>
<td><strong>buaku to-toka</strong></td>
</tr>
<tr>
<td></td>
<td>1 REL-‘perch’</td>
<td>2 REL-‘perch’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tinputz</strong></td>
<td><strong>tonima mea kukon</strong></td>
<td><strong>tonima mea tana</strong></td>
</tr>
<tr>
<td></td>
<td>(5) LIG 3</td>
<td>(5) LIG 4</td>
</tr>
<tr>
<td><strong>Hahon</strong></td>
<td><strong>taunima mana kukana</strong></td>
<td><strong>taunima mana tana</strong></td>
</tr>
<tr>
<td></td>
<td>(5) LIG 3</td>
<td>(5) LIG 4</td>
</tr>
<tr>
<td><strong>Teop</strong></td>
<td>((taonim\ ea)\ kukan\ to\ toka)</td>
<td>((taonim\ ea)\ dorana\ to\ toka)</td>
</tr>
<tr>
<td></td>
<td>(5) LIG 3 REL ‘perch’</td>
<td>(5) LIG 4 REL ‘perch’</td>
</tr>
<tr>
<td><strong>Teop(^{M&amp;T})</strong></td>
<td><strong>kukan\ to-toka</strong></td>
<td><strong>dorana\ to-toka</strong></td>
</tr>
<tr>
<td></td>
<td>3 REL-‘perch’</td>
<td>4 REL-‘perch’</td>
</tr>
</tbody>
</table>

### 6.5 Details of paraphrase in the NB (non-Oceanic) group

Table 6 presents data for three North Bougainville (non-Oceanic) languages. For Konua, Müller explains that in numerals \(upu\) means ‘cover, over’ (1954:13). Later in his dictionary it is clear that \(upu\) (there \(\ddot{u}pu\)) means ‘cover up like a bandage’ (1954:122). The logic here is quite close to that inferred for nearby Teop. Thus \(tar\text{ia-'}upu\) 6 exhibits the pattern \{INCREMENT ‘cover’\}. The alternative given for 6 \(upu\text{-}hara\text{-}kassio\) can be interpreted as \{‘cover’-HAB-5\}. Müller further explains that the article \((mo)\) or \((bo)\) is usually omitted from simple enumeration but is used with definite objects (Müller 1954:14).

\(^{16}\) I suspected the transcription errors when consulting Mosel and Tieson (2007). This suspicion was kindly confirmed by Ulrike Mosel in person at COOL-7, July 2007.

\(^{17}\) Teop\(^{M&T}\) data from Mosel and Thiesen (2007).
Just south of Konua there is a data gap across the island. Counting words past 5 or 6 are not available for Keriaka, or for the variety of Piva spoken in neighbouring Amun village or for Papapana on the east coast. Presumably, Keriaka does paraphrase, as may be inferred from Firchow’s observation about all the non-Austronesian languages being ‘quinary’. The documentation for Rotokas is much better, but it does not clarify the logic. Rotokas vatara seems to mean ‘more (than implied 5)’. The dictionary gives no meaning for vatara by itself. Rotokas does not exactly match Banoni: the meaning of bena is matched better by Rotokas iviro ‘cross over (stream or gorge)’ (Firchow et al. 1973:291).

Table 6: Konua, Rotokas, and Eivo paraphrase \{6,7,8,9\} with morpheme guesses

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Konua</td>
<td>taria-'upu/upu-hara-kassio</td>
<td>(bo)-suku-'upu</td>
</tr>
<tr>
<td></td>
<td>1-‘cover’/‘cover’-HAB-5</td>
<td>(ART)-2-‘cover’</td>
</tr>
<tr>
<td>Rotokas</td>
<td>kai vatara</td>
<td>erao vatara</td>
</tr>
<tr>
<td></td>
<td>1 ‘more’</td>
<td>2 ‘more’</td>
</tr>
<tr>
<td>Eivo</td>
<td>apuketoli</td>
<td>meša-apuketoli</td>
</tr>
<tr>
<td></td>
<td>‘more’</td>
<td>2 ‘more’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Konua</td>
<td>(mo)-heruba-'upu</td>
<td>(mo)-lesiura-'upu</td>
</tr>
<tr>
<td></td>
<td>(ART)-3-‘cover’</td>
<td>(ART)-4-‘cover’</td>
</tr>
<tr>
<td>Rotokas</td>
<td>vopeva vatara</td>
<td>voresiura vatara</td>
</tr>
<tr>
<td></td>
<td>3 ‘more’</td>
<td>4 ‘more’</td>
</tr>
<tr>
<td>Eivo</td>
<td>penumma apuketodi</td>
<td>detšiona apuketodi</td>
</tr>
<tr>
<td></td>
<td>3 ‘more’</td>
<td>4 ‘more’</td>
</tr>
</tbody>
</table>

Documentation for Eivo is limited to the counting words themselves: 6 apuketoli, 7 mesapuketoli, 8 penumma-apuketoli; 9 detsiona-apuketoli.\(^{18}\) The phonology of Eivo is still unstudied but there are hints that, as in the famous dialect of Rotokas (Robinson 2006), voiced consonants /b, d, g/ vary from stop to continuant to nasal, in particular [b]-[v]-[m] are allophones of /b/ as are [d]-[r]-[n] of /d/. Thus, the first two syllables of Eivo besatuaare 2 may be phonemically identical to the beginning of mesapuketoli 7. In Rotokas, [s] and [ts] are allophones of /t/ before [i]. Thus, Rotokas voresiura 4 and Eivo ne’siura 4 from Allen and Hurd’s survey (1963) and Eivo detsiuna 4 from Oliver (n.d.) may actually all have the same phonemic sequence /detiuda/. The Eivo trailing element apuketoli may be glossed as ‘more’. However since, like Banoni bena 6, Eivo apuketoli 6 lacks an overt increment implying first a value of six and then of five, the actual meaning and logic remains hidden.

Overall, the structure of NB counting with the increment preceding the logic element is the reverse of BP. More significantly, the logic is explicitly different in Konua ‘on top’ versus ‘cross’ and probably different in Rotokas. Eivo is like BP in that the numeric element is missing from 6, but the similarity may end there. The basic pattern inferred for this family is simply \{INCREMENT ‘more’\}.

---

\(^{18}\) These Eivo words come from field notes collected by Douglas Oliver in 1938–39 (Oliver n.d.). He kindly shared them with me. I hand-copied them in 1973. I cannot now check the data, but detsiona apuketoli 9 may contain a transcription error of apuketoli for apuketoli.
6.6 Details of paraphrase in the SB (non-Oceanic) group

SB languages (see Table 7) have much more complex counting words than any other languages around. They have dozens of noun classes signalled by inflecting the counting words. Moreover, some numbers include marking for dual, paucal, and plural. As if that weren’t enough, each of the languages has its own logic to paraphrase 6,7,8,9.

Table 7: Details for counting houses \{6,7,8,9,10\} in SB languages

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motuna</td>
<td>na-pa-ngi na-rang 1-CLF-ERG 1-CLF:10</td>
<td>ki-pa-ke na-rang 2-CLF-DU.ERG 1-CLF:10</td>
</tr>
<tr>
<td>Buin</td>
<td>tugi-pa-i 6-CLF-PL</td>
<td>pai-pa-i tuo 3-CLF-PL ‘less’</td>
</tr>
<tr>
<td>Nasioi</td>
<td>pa’noko’ keta na-va-a-taa 5 ‘from’ 1-CLF-‘extra’</td>
<td>pa’noko’ keta ke-va-a-raa-taa 5 ‘from’ 2-CLF-DU-‘extra’</td>
</tr>
<tr>
<td>Nagovisi</td>
<td>na-visi ke nola 1-CLF ‘toward’ 10</td>
<td>ke-visi-la ke nola 2-CLF-DU ‘toward’ 10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motuna</td>
<td>pee-pa-i-nga na-rang 3-CLF-PAU-ERG 1-CLF:10</td>
<td>kori-kang-ngi na-rang 4-CLF-ERG 1-CLF:10</td>
<td>na-rang 1-CLF:10</td>
</tr>
<tr>
<td>Buin</td>
<td>ki-porako tuo 2-CLF-DU ‘less’</td>
<td>kampuro</td>
<td>kii-puro 10</td>
</tr>
<tr>
<td>Nasioi</td>
<td>pa’noko’ keta bee-va-a-ri-taa 5 ‘from’ 3-CLF-PL-‘extra’</td>
<td>pa’noko’ keta ke-va-a-raa-taa 5 ‘from’ 4-CLF-PL-‘extra’</td>
<td>kivora 10</td>
</tr>
<tr>
<td>Nagovisi</td>
<td>we-visi-li ke nola 3-CLF-PL ‘toward’ 10</td>
<td>kale-visi-la ke nola 4-CLF-PL ‘toward’ 10</td>
<td>nola 10</td>
</tr>
</tbody>
</table>

Motuna is the best analysed and the most complicated. The details here come from Masayuki Onishi’s 1994 doctoral grammar of Motuna. The language has some fifty class markers that occur with numerals, other quantifiers and a few other words, e.g. -na’ is used when counting taro, -wui when counting other root vegetables, -meng when counting cutting tools, -pa when counting shelters, and so on. Various parts of the words change depending on the surrounding sounds. The number 2 by itself and in the compound for 7 is inflected for dual. The numbers 3 and 4 by themselves and in the compounds for 8 and 9 are inflected for paucal. For most classes, 4 is represented by kori-. For a few classes, 4 has the form kori-kang- where -kang- replaces the classifier, neutralising the semantic information of a class marker. In Table 7, the Motuna -pa- signals ‘shelters’, but for 9 in place of -pa- there is -kang-, which could also replace various other class markers. Angumuka [lit. ‘hand-like’\(^\text{19}\)] is used unadorned with almost all classes to mean 5. For

\(^{19}\) Gloss from Oliver (1955:100). Could this be Austronesian influence? Probably not. If so, not from Banoni, because irregular changes obscure earlier unity, e.g. *lima ‘hand’ and 5 are now numa and ghinima respectively.
humans, there is an alternate form puu-noruu (one of three such exceptional classes). The form na-rang is the same for most classes. In fact, -rang is a classifier for tens: na-rang thus means ‘one ten.’

The Buin system as presented by Laycock (2003) has a very similar set of noun classes but with simpler structures for the counting words. The forms in Table 7 are used to count houses or buildings. The -pa- appears to be cognate with the Motuna -pa-. The form -porako may some how be inflected for dual. Laycock does not analyse the dual, plural inflection among numerals, but his paradigms for kin terms show singular, dual and plural forms. Across kin paradigms, the duals always end in o and the plurals end in i [except u once, and o twice]; across number sets 2 always ends in o and 3,4,5,6 end in i one third of the time, all of which suggests dual and plural inflection in the Buin counting words. The Buin system has a stem tugi-6 that patterns like the first five numbers with class and plural suffixing. The words for 9 and 10 are uninflected for most classes, but appear with a suffix with some classes, e.g. kampuro-i ‘9 fathoms’, kipuro-i ‘10 fathoms’. What sets Buin apart most is the fact that only 7 and 8 are compounded from smaller numbers and the smaller numbers are used subtractively. Banoni is similar in that only 7 and 8 are compounded from smaller numbers, but how it is done could hardly be more different. Indeed, the subtractive logic is rare in the Pacific.

The Nasioi system has a quite similar system of noun classes marked on the counting words and on some other word classes (Hurd 1977). The classifier -vaa- (presumably cognate with Motuna and Buin -pa-) signals that houses are being counted. The increment is inflected for dual or plural as appropriate. The -taa ‘extra’ on the end makes the pa’noko’keta somewhat redundant. Indeed, as in Teop, the explicit reference to 5 can be omitted, e.g. ke-vaa-raa-taa means ‘seven houses’ from ke-vaa-raa ‘two houses’.

Nagovisi is spoken on the western slopes of the same mountains that are peopled by Nasioi speakers on the eastern slopes. The languages are quite closely related, but with some differences, as evidenced by the counting words (present data based on Lean 1991). The classifier for houses is different: visi vs. vaa. The logic {‘toward’ 10} vs. {‘from’ 5} is different.

The greater unity of the four SB languages is cross cut by cognate pairs for 10: Nagovisi nola 10 appears cognate with Motuna na-rang, while Buin kiipuro and Nasioi kivora seem to form a different cognate pair. Further, SB languages overall show surprising differences; for example in paraphrasing, Nagovisi like Motuna counts up toward 10, Buin counts down from 10 and Nasioi counts up from 5. In general, the quite similar languages of this family have quite different counting word patterns, as shown below.

- Motuna 6,7,8,9: {INCREMENT-CLASSIFIER-ERGATIVE.(0/DUAL/PAUCAL)
  1-CLASSIFIER:10}
- Buin 7,8: {DECREMENT-CLASSIFIER.PLURAL ‘less’ }
- Nasioi 6,7,8,9: {5 ‘from’ INCREMENT-CLASSIFIER.(0/DUAL/PLURAL)-‘extra’}
- Nagovisi 6,7,8,9: {INCREMENT ‘toward’ 10:CLASSIFIER}

The SB counting words themselves are substantially more complex than BP words and are unlikely to have motivated change in Banoni. Most notable are the noun classes. When I pressed for classifiers in Banoni counting, the most I found was partial reduplication to indicate counting round objects. However, even this was not consistent: reduplication was
only expressed on 1,2,3 then either there was no change (4, 8) or the causative prefix\textsuperscript{20} was used (5,6,7,9,10): ka-kadaken, to-toom, da-dapisa, tovatsi, va-ghinima, va-bena, va-benatom, bena kapisa, va-visa, va-manogha (Lincoln 1976c:98).

7 Conclusions

The five steps of this paper justify the following observations:

- Banoni has several apparent prefixes on counting words, but given the ebb and flow of such accretions Banoni represents just an eddy of such flow;
- Banoni adopted pisa 3, probably through direct or indirect contact with languages on Lauru;
- Banoni adopted manonya 10, probably through direct or indirect contact with languages on Lauru;
- Motivation for the pattern of paraphrase 6,7,8 involving bena ‘cross’ is unclear (no other languages quite match the BP system); and
- Banoni and the BP system alone paraphrase only 6,7,8.

8 Discussion

The expectation when approaching this study was that Banoni (and Piva) created a unique system of paraphrase independently, free of any neighbouring language influence. There are two problems that stall confirmation of that hypothesis. First, not all of the neighbouring languages are available for comparison; not all of those available are sufficiently documented to decide how dissimilar the systems are. Secondly, even in languages with complete data, it is difficult to determine the degree of similarity that is sufficient to imply contact and to exclude chance or independent change. The close relation of the four languages of SB family seems contradicted by the surprisingly diverse structure of their paraphrase. The subtractive logic in paraphrase is rare, perhaps occurring in only four spots in the Pacific: Yap, Manus, Buin, and Santa Cruz (Bradshaw 1995; Lincoln 1978). There may be a linguistic Manus-Yap connection (Ross 1988:326–329), but given the distances involved chance remains a plausible explanation. At the other end of the distance scale, Lean’s Tinputz and Blackwood’s Kurtatchi belong to the same language or at least the Hahon-Teop-Tinputz group, but Tinputz paraphrases from 6 and Kurtatchi starts at 11.

The foundation of comparative linguistics is that for most words the relation between sound (among hundreds of possibilities) and meaning (among tens of thousands of possibilities) is arbitrary,\textsuperscript{21} so that when one finds similar sounding words with similar meanings in two languages, one expects at least the possibility of shared ancestry or some other type of language contact. Thus, Banoni tovatsi 4 is considered to be derived from POc *pati 4, and the similarity of Ririo pisa 3 and Banoni dapisa 3 implies at some point these people were talking to each other or to the same set of intermediaries. Counting in

\textsuperscript{20} The va- causative prefix is used more consistently to express ordinals after mama ‘first’, va-toom ‘second’, va-dapisa ‘third’, va-tovatsi ‘fourth’ and so on (Lincoln 1976c:98).

\textsuperscript{21} For example, Banoni daame does not have to mean ‘good’; indeed much the same sound sequence in Japanese dame means ‘bad’.


one language from 1 to 10 usually involves words for which the arbitrary sound-meaning relation is shared and regular, as it does for other words. Most folk use ten separate words and have a decimal system, start paraphrasing at 11 to derive the rest of the system up to one hundred and then add a new word for one hundred to be used to build words on up to one thousand and so on. Some of the languages in this study, start paraphrasing at 6 (Buin at 7), but there are not so very many ways that can be done, and most seem to have been tried in the data examined here.

The languages belong to three radically different groups: Oceanic, North Bougainville and South Bougainville. Similarities cross group boundaries and radical differences occur between closely related varieties. Little significance can be attached to the position at which the languages start paraphrasing, at 6 or at 11. In his discussion of number systems in Oceanic languages, Blust (2005) suggests that the paraphrase of 6,7,8,9 in parts of Vanuatu and most if not all of New Caledonia imply that there was a Papuan substratum in remote Oceania. Blust adduces further evidence that the current understanding of the archeology is wrong, but one goal of this article is to caution scholars not to count on substrata for such support. While a Papuan substratum may be necessary to make an Oceanic language start paraphrasing at 6 rather than 11, a Papuan substratum is not sufficient explanation for this. In Bougainville, Uruava continued to use Proto Oceanic derived words to count to 10 until the language finally lost all its potential speakers to Nasiol in the 1970s. In the Solomons from Alu to Makira, all the languages use ten words to count to 10, even the four surviving non-Austronesian languages. Presumably the Papuan languages were more widespread in the area in the past without causing the current languages to start paraphrasing at 6.

This is my third attempt to find detailed empirical evidence of Papuan influence. In each case, substratum influence has remained elusive. Lincoln (1976b) attempted to find evidence in neighbouring Papuan languages that might explain the ways in which Piva differed from Banoni. However, far from being Papuanised, Piva turned out to retain more Proto Oceanic features than Banoni did. Lincoln (1978) sought to show that Reef Santa Cruz languages developed without Papuan influence. Oceanic traits were found in those languages; however, until more external Oceanic cognates are found and analysed, Papuan strata cannot be ruled out. The current study has sought evidence of Papuan influence that might account for the development of the modern Banoni counting system. Though surprisingly varied, the neighbouring systems offer no obvious model for Banoni, suggesting that the diversity is largely driven by internal creativity rather than by substrata influences.
## Appendix

**Table A1:** Counting words for NWS linkage with POc forms from Lean (1991), Lincoln (1976b), Tryon and Hackman (1983)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>POc</td>
<td>*sa</td>
<td>*rua</td>
<td>*tolu</td>
<td>*pati</td>
<td>*lima</td>
<td>*onom</td>
</tr>
<tr>
<td>Nehan</td>
<td>kas</td>
<td>torik</td>
<td>tawon</td>
<td>tovat</td>
<td>tolim</td>
<td>manom</td>
</tr>
<tr>
<td>Halia</td>
<td>toa</td>
<td>tol</td>
<td>topisa</td>
<td>tohats</td>
<td>tolima</td>
<td>tunomo</td>
</tr>
<tr>
<td>Solos</td>
<td>mes</td>
<td>huenu</td>
<td>huepis</td>
<td>nina</td>
<td>talinng</td>
<td>tonom</td>
</tr>
<tr>
<td>Petats</td>
<td>sea</td>
<td>hualu</td>
<td>hopis</td>
<td>nina</td>
<td>tolim</td>
<td>tonom</td>
</tr>
<tr>
<td>Saposa</td>
<td>isen</td>
<td>fuan</td>
<td>fopis</td>
<td>fats</td>
<td>ngim</td>
<td>aunom</td>
</tr>
<tr>
<td>Hahon</td>
<td>paia</td>
<td>buaku</td>
<td>kakana</td>
<td>tana</td>
<td>taunima</td>
<td>taunima mana paia</td>
</tr>
<tr>
<td>Tinpuzt</td>
<td>paeh</td>
<td>bok</td>
<td>kakon</td>
<td>tana</td>
<td>tonim</td>
<td>tonima mea paeh</td>
</tr>
<tr>
<td>Teop</td>
<td>peha</td>
<td>buaku</td>
<td>kakan</td>
<td>dorana</td>
<td>taonim</td>
<td>(5+) peha to toka</td>
</tr>
<tr>
<td>Papapana</td>
<td>na’?aria</td>
<td>nu’ata?</td>
<td>tau’tonu</td>
<td>tau’vasi</td>
<td>papai’tau nima</td>
<td></td>
</tr>
<tr>
<td>Uruava</td>
<td>kaa</td>
<td>rua</td>
<td>toru</td>
<td>uasi</td>
<td>rima</td>
<td>onomo</td>
</tr>
<tr>
<td>Torau</td>
<td>kasa</td>
<td>erua</td>
<td>episa</td>
<td>evati</td>
<td>nima</td>
<td>olomo</td>
</tr>
<tr>
<td>Mono-Alu</td>
<td>kaala</td>
<td>ela</td>
<td>episa</td>
<td>ehati</td>
<td>lima</td>
<td>onomo</td>
</tr>
<tr>
<td>Banoni</td>
<td>kadaken</td>
<td>toom</td>
<td>dapisa</td>
<td>tovatsi</td>
<td>ghinima</td>
<td>bena</td>
</tr>
<tr>
<td>Piva</td>
<td>kadakeni</td>
<td>tonua</td>
<td>topisa</td>
<td>evatsi</td>
<td>nima</td>
<td>ebena</td>
</tr>
<tr>
<td>Vaghua</td>
<td>kala</td>
<td>karua</td>
<td>kapa</td>
<td>kavats</td>
<td>kaloma</td>
<td>kaka</td>
</tr>
<tr>
<td>Ghone</td>
<td>kala/kameka</td>
<td>karua</td>
<td>kapisa</td>
<td>kavasi</td>
<td>kalima</td>
<td>kaka</td>
</tr>
<tr>
<td>Varisi</td>
<td>kala</td>
<td>karua</td>
<td>kapisa</td>
<td>kavasi</td>
<td>kalima</td>
<td>kaka</td>
</tr>
<tr>
<td>Ririo</td>
<td>kik</td>
<td>ker</td>
<td>pisa</td>
<td>vets</td>
<td>lima</td>
<td>voso</td>
</tr>
<tr>
<td>Bambatana</td>
<td>koe</td>
<td>kere</td>
<td>tulu</td>
<td>vati</td>
<td>lima</td>
<td>voso</td>
</tr>
<tr>
<td>Sengga</td>
<td>kike</td>
<td>keri</td>
<td>tulu</td>
<td>vati</td>
<td>limma</td>
<td>soso</td>
</tr>
<tr>
<td>Lōmaumbi</td>
<td>kike</td>
<td>keri</td>
<td>passa</td>
<td>vati</td>
<td>limma</td>
<td>soso</td>
</tr>
<tr>
<td>Avasō</td>
<td>korkki</td>
<td>keri</td>
<td>passa</td>
<td>vati</td>
<td>limma</td>
<td>soso</td>
</tr>
<tr>
<td>Ghanongga</td>
<td>maka</td>
<td>kiri</td>
<td>kue</td>
<td>ma^de</td>
<td>lima</td>
<td>voso</td>
</tr>
<tr>
<td>Simbo</td>
<td>kame</td>
<td>karu</td>
<td>kue</td>
<td>ma^de</td>
<td>lima</td>
<td>voso</td>
</tr>
<tr>
<td>Nduke</td>
<td>keka</td>
<td>kori</td>
<td>kuhe</td>
<td>ka^de</td>
<td>lima</td>
<td>voso</td>
</tr>
<tr>
<td>Ughele</td>
<td>taza/kaike</td>
<td>karua</td>
<td>neta</td>
<td>ma^de</td>
<td>lima</td>
<td>soso</td>
</tr>
<tr>
<td>Roviana</td>
<td>taza/keke</td>
<td>karua</td>
<td>neta</td>
<td>ma^de</td>
<td>lima</td>
<td>soso</td>
</tr>
<tr>
<td>Kusaghe</td>
<td>keke</td>
<td>karua</td>
<td>neta</td>
<td>ma^de</td>
<td>lima</td>
<td>soso</td>
</tr>
<tr>
<td>Hoava</td>
<td>keke</td>
<td>karua</td>
<td>hike</td>
<td>ma^de</td>
<td>lima</td>
<td>soso</td>
</tr>
<tr>
<td>Bareke</td>
<td>meka</td>
<td>karua</td>
<td>hike</td>
<td>ka^buto</td>
<td>kalima</td>
<td>kaka</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOc</td>
<td>*pitu</td>
<td>*walu</td>
<td>*siwa</td>
<td>*sapapulu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nehan</td>
<td>movit</td>
<td>tuwal</td>
<td>lusio</td>
<td>hangaul</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halia</td>
<td>tohit</td>
<td>toal</td>
<td>losi</td>
<td>maloto</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solos</td>
<td>tohit</td>
<td>tuen</td>
<td>tosia</td>
<td>manot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Petats</td>
<td>tohit</td>
<td>towal</td>
<td>tosia</td>
<td>malot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saposa</td>
<td>fits</td>
<td>jian</td>
<td>sia</td>
<td>sajuna</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hahon</td>
<td>taunima mana buaku toka</td>
<td>taunima mana kunakan toka</td>
<td>taunima mana tana</td>
<td>savun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tinputz</td>
<td>tonima mea bok</td>
<td>tonima mea kukon toka</td>
<td>tonima mea tana</td>
<td>havun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teop</td>
<td>(5+)* buaku to toka</td>
<td>(5+)* kakan to toka</td>
<td>(5+)* dorana to toka</td>
<td>savun</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papapana</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uruava</td>
<td>uitu</td>
<td>aru</td>
<td>ia</td>
<td>avuru</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Torau</td>
<td>itu</td>
<td>anu</td>
<td>sia</td>
<td>saunu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mono-Alu</td>
<td>hitu</td>
<td>aalu</td>
<td>ulia</td>
<td>laahulu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banoni</td>
<td>bena toom</td>
<td>bena kapisa</td>
<td>visa</td>
<td>monogha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piva</td>
<td>bena tonua</td>
<td>bena topisa</td>
<td>sia</td>
<td>monogha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaghua</td>
<td>kavuts</td>
<td>kazal</td>
<td>kaxasa</td>
<td>naval</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghone</td>
<td>kavitu</td>
<td>kazalu</td>
<td>kajia</td>
<td>manawa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varisi</td>
<td>kavitu</td>
<td>kazalu</td>
<td>kaija</td>
<td>manγa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ririo</td>
<td>ziuts</td>
<td>zai</td>
<td>zia</td>
<td>mano</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bambatana</td>
<td>vitu</td>
<td>viu</td>
<td>ia</td>
<td>meina</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sengga</td>
<td>vettu</td>
<td>viu</td>
<td>zia</td>
<td>mano</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lömaumbi</td>
<td>veto</td>
<td>viu</td>
<td>ia</td>
<td>meina</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avaso</td>
<td>vittu</td>
<td>viu</td>
<td>ia</td>
<td>meina</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghanongga</td>
<td>vitu</td>
<td>vesu</td>
<td>sia</td>
<td>manγa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simbo</td>
<td>vitu</td>
<td>kalu</td>
<td>sia</td>
<td>manγa putea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nduke</td>
<td>vitu</td>
<td>vesu</td>
<td>sia</td>
<td>manγa putea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ughele</td>
<td>njuapa</td>
<td>alu</td>
<td>niki</td>
<td>manγe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roviana</td>
<td>zuapa</td>
<td>vesu</td>
<td>sia</td>
<td>manγe putea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kusaghe</td>
<td>zuapa</td>
<td>vesu</td>
<td>sia</td>
<td>manγe putea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoava</td>
<td>zuapa</td>
<td>vesu</td>
<td>sia</td>
<td>manγe putea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bareke</td>
<td>kanjuapa</td>
<td>kavesu</td>
<td>kasia</td>
<td>nu'guru putea</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) (5 +)* is an abbreviation for the optional elements (taunim ea) in Teop.

(b) Bambatana, Sengga [Sisīqə], Lōmaumbi, Avaso are all varieties of East Choiseul for Tryon and Hackman (1983), but see also Lynch et al. (2002:456).
<table>
<thead>
<tr>
<th>Language</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NB languages:</strong> data from Lean (1991); Konia\textsuperscript{M} Müller (1954); Rotokas\textsuperscript{F} Fireholt, et al. (1973); Eivo\textsuperscript{O} and Keriaka\textsuperscript{O}: Oliver (n.d.).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Konia\textsuperscript{M}</td>
<td>taria</td>
<td>suku</td>
<td>heruba</td>
<td>lestura</td>
<td>urou or bokassio</td>
<td>taria’upu or upuharakassio</td>
<td>suku’upu</td>
<td>heruba’upu</td>
<td>lesiura’upu</td>
<td>kalurua or aretevua</td>
</tr>
<tr>
<td>Rotokas\textsuperscript{F}</td>
<td>katai</td>
<td>erao</td>
<td>voceva</td>
<td>voresiura</td>
<td>vovavae</td>
<td>katai vatara</td>
<td>erao vatara</td>
<td>voceva vatara</td>
<td>voresiura vatara</td>
<td>katai taa</td>
</tr>
<tr>
<td>Eivo</td>
<td>kantepato</td>
<td>bestia’toar</td>
<td>pen’uma</td>
<td>ne’siuna</td>
<td>a’kamoko</td>
<td>apukapairan</td>
<td>apukato</td>
<td>mo’siapukato</td>
<td>ponuma apukato</td>
<td>dotsona apukato</td>
</tr>
<tr>
<td>Keriaka</td>
<td>lo ovesi</td>
<td>pe’numa</td>
<td>tu’biri</td>
<td>mo’nisuka</td>
<td>’mahai</td>
<td>mo’ra’Juapa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keriaka\textsuperscript{O}</td>
<td>raf</td>
<td>penuma</td>
<td>hosi</td>
<td>nisuma</td>
<td>mafi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SB languages:</strong> data from Lean (1991).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasoisi</td>
<td>narung</td>
<td>kenanka</td>
<td>benaumo</td>
<td>karemuamo</td>
<td>panoko</td>
<td>panoko keta</td>
<td>narung taa</td>
<td>narung taa</td>
<td>karoano maa</td>
<td>karoano maa</td>
</tr>
<tr>
<td>Nagovisi</td>
<td>nau</td>
<td>keka</td>
<td>wekogo</td>
<td>kalekago</td>
<td>panoko</td>
<td>narung taa</td>
<td>karoano maa</td>
<td>karoano maa</td>
<td>karoano maa</td>
<td>karoano maa</td>
</tr>
<tr>
<td>Ipu</td>
<td>nonumoi</td>
<td>keitako</td>
<td>paigami</td>
<td>kongkongi</td>
<td>upugami</td>
<td>tigigami</td>
<td>paigami tno</td>
<td>keitako tno</td>
<td>keitako tno</td>
<td>keitako tno</td>
</tr>
<tr>
<td>Motuna</td>
<td>mo’o</td>
<td>ki’ko</td>
<td>pekangi</td>
<td>korikangi</td>
<td>angumanka</td>
<td>no’oki narung</td>
<td>narung taa</td>
<td>narung taa</td>
<td>narung taa</td>
<td>narung taa</td>
</tr>
<tr>
<td><strong>CS languages:</strong> data from Tryon and Hackman (1983).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biliua</td>
<td>xu’mde</td>
<td>xu’mga</td>
<td>zouke</td>
<td>ariku</td>
<td>sile</td>
<td>varimu’ja</td>
<td>sikera</td>
<td>siotosu</td>
<td>siakava</td>
<td>tesi</td>
</tr>
<tr>
<td>Baniata</td>
<td>ru’u</td>
<td>ru’u</td>
<td>heke</td>
<td>awo</td>
<td>soud</td>
<td>ru’u’bi</td>
<td>sikua</td>
<td>siotosu</td>
<td>siakava</td>
<td>tesi</td>
</tr>
<tr>
<td>Loku</td>
<td>ru’u</td>
<td>ru’u</td>
<td>heke</td>
<td>awo</td>
<td>soud</td>
<td>ru’u’bi</td>
<td>sikua</td>
<td>siotosu</td>
<td>siakava</td>
<td>tesi</td>
</tr>
<tr>
<td>Kazukuru</td>
<td>xu’u</td>
<td>xu’u</td>
<td>heke</td>
<td>awo</td>
<td>soud</td>
<td>ru’u’bi</td>
<td>sikua</td>
<td>siotosu</td>
<td>siakava</td>
<td>tesi</td>
</tr>
<tr>
<td>Lavukaleve</td>
<td>dom’tetel</td>
<td>lemal/lela</td>
<td>ena</td>
<td>nana</td>
<td>see</td>
<td>ce</td>
<td>roa</td>
<td>sevi</td>
<td>savu</td>
<td>kanam</td>
</tr>
<tr>
<td>Savosavo</td>
<td>ela’pa</td>
<td>e’do</td>
<td>iya</td>
<td>aya</td>
<td>ara</td>
<td>poyo</td>
<td>poyo</td>
<td>poyo</td>
<td>poyo</td>
<td>poyo</td>
</tr>
</tbody>
</table>

* The forms given in this table for SB languages differ from those given in Table 7 above, because the former use the general classifier form whereas the latter are the forms used for counting houses.
References


19 Lexical history in the Northwest Solomonic languages: evidence for two waves of Oceanic settlement in Bougainville and the northwest Solomons

MALCOLM ROSS

1 Introduction

Sheppard, Walter and Roga (this volume) summarise archaeological evidence for the settlement history of the northwest Solomons (Mono and Alu, Choiseul, the New Georgia group and Santa Isabel). They refer to a proposal based on circumstantial linguistic evidence that I advanced twenty years ago to the effect that there had been two waves of Oceanic settlement in the northwest Solomons (Ross 1988:382–386). The proposal has not been widely accepted among Oceanist linguists, and this paper seeks to offer, among other things, more direct linguistic support for the two-wave proposal.

Andrew Pawley, in whose honour this volume is published, has a longstanding interest in the linguistic history of the Solomon Islands, and especially of the Southeast Solomonic (SES) languages (Pawley 2009). I shall say almost nothing here about the SES group, but the history of their immediate neighbours in the Northwest Solomonic (NWS) group must provide at least one piece in the SES historical jigsaw.

---

1 I owe a very considerable debt of gratitude to Andy. It was one of his papers (Pawley 1975) which inspired my first foray into historical linguistic research, emulating his title (Ross 1977), and he was one of those who encouraged me in my late entry into the field. His arrival at the Australian National University came shortly after my appointment there, and he has been a mentor and friend ever since. I am also grateful to him and to Bethwyn Evans for comments on earlier drafts of this paper.

The reader may note that the research leading to the present paper and to Pawley (2009) overlaps. This is because unbeknownst to each other we were doing this work at the same time. Thanks to the delay in publishing the present volume, I have nonetheless been able to refer here to Pawley (2009).
2 Northwest Solomonic: languages and abbreviations

NWS languages fall into six groups:²

(1) a. North Bougainville (NBv):³ Nehan, Solos, Petats, Haku, Selau, Taiof, Teop
   b. Banoni–Piva (BP): Banoni, Piva
   c. Mono-Torau (MT): Mono, Torau, Uruava (extinct)
   d. Choiseul (Ch):
      i. West Choiseul (WCh): Vaghua, Varisi, Ririo
      ii. East Choiseul (ECh): Ririo, Babatana, Sisingga
   e. New Georgia (NGe):
      i. West New Georgia (WNGe): Simbo, Lungga, Ghanongga, Nduke
      ii. East New Georgia (ENGe): Nduke, Roviana, Ughele, Kusaghe, Hoava, Marovo (Mvo), Vangunu
   f. Santa Isabel (Is): (Kia, Kokota, Laghu, Kilokaka (Zazao), Blablanga, Ghove, Maringe (Mge)

The placement of Ririo in both West and East Choiseul and of Nduke in both West and East New Georgia is not an error, but a mark of the fact that each of these languages provides a transition between the two parts of its respective group. Abbreviations for language names consist of the first three letters of the name except in the cases of Marovo (Mvo) and Maringe (Mge). Abbreviations of protolanguage names are formed by placing lower-case ‘p’ before the abbreviation for the group name, for example, pNWS for Proto Northwest Solomonic.

The close relatedness of Mono and Torau appears surprising in the light of their present locations, but until less than two centuries ago Torau speakers were evidently located in the extreme southeast of Bougainville, just across the water from Mono (Terrell and Irwin 1972).

3 Northwest Solomonic: genealogy

The NWS group belongs to the larger Western Oceanic linkage (a linkage is a group of languages that has emerged from an earlier dialect network and is paraphyletic, i.e. lacks an ancestor which has no other descendants). Western Oceanic comprises those languages of the Oceanic subgroup of Austronesian that are located on the north coast of West Papua, in Papua New Guinea excluding the Admiralties and Mussau, and in the northwest Solomons (Ross 1988:386–389; Lynch, Ross and Crowley 2002:96).

Proto Oceanic (pOc) must have been spoken in an area which included the Admiralties, Mussau Island, and parts of New Britain and New Ireland and their offshore islands (Pawley 2008). Its speakers were associated with the early phase of the archaeologically recognisable Lapita culture around 1400 BC. They spread early to Mussau Island and the Admiralties and by 1100 BC had made their first south-eastward push reaching the Reef and Santa Cruz Islands and northern Vanuatu, whence they moved eastward to settle Fiji.

² The languages listed here are those referred to in this paper. For a full listing see Ross (1988:217) or, with a different nomenclature, Tryon and Hackman (1983). Maps showing the locations of languages can be found in Ross (1988), Ross, Pawley and Osmond (1998) or Ross, Pawley and Osmond (2003).
³ Labelled ‘Nehan/North Bougainville’ in Ross (1988).
Polynesia and Micronesia and southward to settle the Vanuatu archipelago, the Loyalties and New Caledonia.⁴

These expansions brought about the break-up of Proto Oceanic, and Western Oceanic appears to have diversified out of a dialect network that was left behind in New Britain and New Ireland. However, before it diversified, certain innovations took place in the network that were not shared by the speakers in the Admiralties, Mussau or the first south-eastward push (Ross 1988:382–383; Lynch, Ross and Crowley 2002:101). In the course of its diversification, Western Oceanic split into two or three separate networks, with a major division at the Willaumez Peninsula on the north coast of New Britain.⁵ The network to the east of the peninsula has been dubbed the Meso-Melanesian (MM) linkage (Ross 1988:257, 423:fn.98).⁶ The linkage has a complex internal structure, in outline as follows:⁷

(2) Meso-Melanesian linkage

Bali, Vitu

Willaumez linkage

New Ireland/Northwest Solomonic linkage

Tungag–Nalik family

Tabar linkage

Madak linkage

Tomoip

St George linkage

South New Ireland languages

Northwest Solomonic family

The NWS family, then, is a portion of the St George linkage and is coordinate within it with a number of South New Ireland languages. The latter appear to form seven groups, each coordinate with NWS, but the detailed history of these relationships is complicated (Ross 1988:258, 306–314, 1997). NWS is labelled a family because it resulted from the dispersal of speakers of a single language, pNWS. We can be confident of this because NWS languages reflect certain innovations not found in other MM languages. They were (Ross 1988:218, 247–249):

(3) a. pOc *w was lost.

b. A vowel was added after a pOc final consonant. This vowel echoed the vowel before the final consonant. For example, pOc *boRok ‘pig’ > pNWS *boroɣo.

c. Following (3b), pOc word-final *-q became pNWS *-k (elsewhere *q became pNWS *ɣ, merging with pOc *k).⁸

---


⁵ I argued in Ross (1988:382) that this was the location of the pOc homeland. Pawley (2008) rightly points out that the evidence says nothing directly about pOc, only about Western Oceanic.

⁶ Regarding its characterisation as a linkage, see Lynch, Ross and Crowley (2002:101).

⁷ Adapted from Lynch, Ross and Crowley (2002:101), where I replaced the clumsy ‘South New Ireland/Northwest Solomonic linkage’ of Ross (1988:258) with the ‘St George linkage’, after Cape St George (the southernmost tip of New Ireland) and the St George’s Channel (between southern New Ireland and the Gazelle Peninsula of New Britain). Language names are in italics.

⁸ In the orthography of Ross (1988) pNWS *ɣ was shown as pNWS *q.
d. The pOc first person singular free pronoun *[i]au acquired an accreted *r-, becoming *[r][i]au.

With regard to (3d), accreted *r- is also mostly reflected on the other first and second person pronouns in the NBv, BP, MT and Ch groups, but not in NGe or Is. This accretion was probably the outcome of major changes in clause order that occurred in pNWS (Ross 1988:228–247), but their discussion lies beyond the scope of this paper.

Certain other NWS innovations relative to pOc had already occurred in MM but are relevant to the interpretation of some of the data given below. There were three mergers: pOc *r and *R merged as early Meso-Melanesian (eMM) *r, pOc *dr and *d as eMM *d, and pOc *s and *c as eMM *s.\(^9\) There were also two apparent splits: pOc *k into eMM *ɣ and *k, pOc *p into eMM *v and *p, but these were almost certainly not unconditioned phonemic splits: instead they were the outcome of borrowing lexical items from languages with lenited *p and *k after lenition had occurred in the borrowing language. NWS innovations are illustrated with supporting data in Ross (1986).

One complex morphosyntactic innovation receives frequent mention in the literature: this is the adoption of what were once possessive noun phrase structures as verb phrase structures (e.g. Ross 1982; Palmer 2002, 2003). This may indeed have been a NWS innovation, but precisely because it is syntactic, it may also have arisen through contact and is thus not a strong candidate for shared inheritance from the protolanguage, pace Ross (1988:249–251). Indeed, it is also reflected in certain SES languages, perhaps as a result of contact.

Each of the six NWS groups except Banoni-Piva is characterised by certain phonological and morphosyntactic innovations relative to pNWS, but in comparison with the innovations characterising NWS as a whole, they are rather insignificant. They are:

(4) a. North Bougainville: pNWS *ɣ is lost; pNWS *u became pNBv *i in certain lexical items; pNBv innovated two noun classes, marked by articles *a and *u; the article is repeated before an attributive adjective (Ross 1988:223, 252–253).

b. Banoni–Piva: none (relationship is obvious by inspection).


d. Choiseul: pNWS *s is lost in some pCh items, retained as pCh *s in others (Ross 1988:224).\(^10\)


f. Santa Isabel: pNWS *s became pIs *f; pNWS medial *-s- often became pIs *-h-; PWNS *tolu ‘three’, *visa ‘how many?’ and *vai ‘where’ became respectively pIs *tilo (expected **tolu), *n-iha (expected **na-fiha) and *hae (expected *fae) (Ross 1988:225).

There are also innovations which are common to the New Georgia and Santa Isabel groups, suggesting that they may have had a short period of exclusively shared history.

---

\(^9\) Since Meso-Melanesian is a linkage, there is no protolanguage from which its members are exclusively descended. I use ‘early Meso-Melanesian’, abbreviated ‘eMM’, to denote the earliest reconstructable stage ancestral to pNWS in which the innovations common to MM languages had occurred.

\(^10\) Ross (1988:224) interpreted reflexes of pNWS *s as pCh *j as a further split. In §7.5 below they are treated as borrowings into pWCh.
However, this hypothesis now appears much weaker than it did in 1988. Proto NWS *rani ‘day’ became *rane (Ross 1988:224). Proto NWS non-first-person-singular free pronouns with accreted *r- do not occur, as noted in connection with (3d): whether this is a shared innovation is not clear. I claimed in Ross (1988:215, 240–247) that Roviana (ENGe) and Maringe (Is) both reflected an innovatory postverbal topic marker *si. Evidence has since become available that it is not reflected in other NGe or Is languages and probably should not be attributed to a shared protolanguage.

4 Papuan languages of north-west Island Melanesia

Crucial to the discussion in this paper is the fact that the Bismarck Archipelago (New Britain, New Ireland and the Admiralty Islands), Bougainville and the northwest Solomons were occupied by Papuan speakers for millennia before the arrival of speakers of Austronesian languages. In this context, ‘Papuan’ simply means ‘not Austronesian’, as there is reasonably good evidence in the form of the surviving Papuan languages of the region that by the time Austronesian speakers arrived, the various groups of Papuan languages had long since lost any indicators of genealogical relationship (assuming that such a relationship once existed) (Ross 2001, 2005; Dunn, Reesink and Terrill 2002; Dunn et al. 2005; Terrill 2002).

Much of this region was settled by speakers of (what we may assume to have been) ancestral Papuan languages during the Pleistocene (Pawley 2007a; Ross forthcoming; Summerhayes 2007). Soon after 19,000 BC, after the Last Glacial Maximum, there was a shift in New Britain from mobile foraging to foraging sedentism (Spriggs 1997:61–65). There are indications that animal and plant species were deliberately imported into New Ireland and Manus. Spriggs (1996, 1997:31–34, 61) interprets this as the beginning of what he calls wildfood production, i.e. the deliberate tending of the forest environment by selective weeding or clearing and by transplanting, without the permanent clearing of the forest which is entailed in agriculture. This situation may have subsisted until the introduction of agriculture by Austronesian speakers, but recent research suggests otherwise. The pre-Lapita distribution of stone pestles and mortars, which appear to have been used for making taro pudding, includes parts of New Britain and New Ireland with conditions appropriate for taro cultivation (Torrence and Swadling 2008), and it is possible that future research will confirm that taro was grown in these islands before the Austronesian arrival.

5 The two-wave proposal

The Solomon Islands are bisected linguistically by a line identified by Tryon and Hackman (1983), which I dubbed the ‘Tryon–Hackman line’ in 1988. It forms the boundary between the NWS and SES languages, which belong to different primary subgroups of Oceanic. The closest relatives of NWS are the languages of southern New Ireland, the next closest the remaining MM languages. SES, on the other hand, has no identifiably close relatives either to the west or the east.

How does one account for this mid-Solomons boundary? Broadly, there are two possible kinds of hypothesis. The first says that the earliest NWS and SES speakers were

---

11 Or Nuclear Oceanic, if one accepts Blust’s division of Oceanic into an Admiralties group and a group containing the rest of Oceanic. See Pawley (2008:57) for nomenclature.
both descended from settlements established during the rapid southeastward expansion before 1100 BC, but that speakers of early Oceanic remained in intense enough contact with each other for a long enough period of time for the innovations which characterise MM to spread through the whole of the early MM dialect network, including pNWS. I find the required intensity of contact over an extended period difficult to believe in. Furthermore, even if I did believe in it, I would still have to explain the hiatus in contact that is reflected in the Tryon–Hackman line.

Because of these difficulties, I continue to prefer a two-wave hypothesis like that put forward in Ross (1988:382–386), which suggests (i) that pSES and the languages of Remote Oceania are outcomes of the expansion before 1100 BC, (ii) that the innovations that characterise MM occurred somewhat later, in the portion of the Western Oceanic linkage to the east of the Willaumez Peninsula of New Britain which extended probably to southern New Ireland and to Tangga and Anir Islands to its east and Nissan to its south;12 (iii) that speakers of a language spoken somewhere in southern New Ireland or on Nissan Island and their descendants moved south-eastward first to Buka and north Bougainville, where their language underwent the innovations that made it pNWS; (iv) that descendants of pNWS speakers occupied coastal enclaves around the rest of Bougainville and then the northwest Solomons. For socioeconomic reasons which perhaps entailed symbiotic relationships with Papuan speakers (cf. Dutton 1994), the NWS south-eastward expansion stopped roughly at the furthest point of much earlier Papuan expansion. At some date after this, NWS speakers came into contact with SES speakers and the Tryon-Hackman line came into being.

The summary by Sheppard et al. (this volume) indicates that some form of two-wave hypothesis enjoys archaeological support, in that work in the NWS region has turned up no signs of early Lapita (i.e. first wave) settlement, despite the presence of such settlements further east. Instead, there is evidence of late Lapita settlement on Buka around 800 BC and in the New Georgia island group around 600 BC, which seems to correlate with the second wave south-eastward spread of MM. However, there is a difference of archaeological opinion as to whether Lapita sailors initially left a sparse population in the NWS region on their way south-eastward (Felgate 2001, 2003, 2007) or leapfrogged it altogether (Sheppard and Walter 2006). Pawley (2008) favours a variant of the former position, inferring that the earliest Oceanic speakers in the Solomons found few of the luxuriant reef systems that were their preferred habitat (and the larger islands occupied by hunter-gatherers) and thus occupied only a few small islands. He is agnostic, however, about whether these early Oceanic speakers eventually expanded to become the ancestors of today’s NWS languages or whether pNWS was brought by a second wave of Oceanic (MM) speakers from around 800 BC which replaced the languages of the very sparse initial Oceanic speaking communities (Pawley 2009:536). Felgate (2007:126–127) favours the latter option, which is supported by the evidence presented below. He sees the first wave of Oceanic speakers in the northwest Solomons as leading a precarious existence which resulted in the displacement of their languages by the later NWS arrivals.

I shall refer to the languages of the first wave of Oceanic speakers (before 1100 BC) in Bougainville and the northwest Solomons as ‘Old Oceanic’ languages.

The MM spread was, one may infer, more gradual and supported by a growing population of speakers, and involved more co-operative relationships with Papuan

---

It is a reasonable inference that when Old Oceanic and MM speakers came into contact, the latter were socially dominant and more populous, and their languages prevailed. As Pawley (2008) points out, however, replacement was not a large-scale process, as Old Oceanic speakers in most cases probably did not live on the larger islands.

It is tempting to believe that pSES was just such an Old Oceanic language. However, there is nothing in common between the phonological innovations manifest in pSES and those attested in the putative Old Oceanic loans reconstructed below.

The linguistic evidence offered for the two-wave proposal has largely been circumstantial. If, however, MM languages do represent a second wave such as I have described, we would at least expect to find loanwords from not only Papuan but also Old Oceanic languages, and perhaps more radical changes resulting from language shift, similar to those found in Madak and Lamusong on New Ireland (Ross 1994). The rest of this paper is a preliminary investigation of the lexicon of languages of the NWS subgroup of MM. It shows that (i) the lexical retention rates of NWS are lower than those of non-NWS Oceanic languages; (ii) there are numerous apparent Papuan loans in NWS, but they cannot readily be sourced; (iii) there are NWS lexical items which appear to be Old Oceanic loans.

6 The retention rates of Northwest Solomonic languages

If the scenario above is roughly correct, we would expect higher retention of pOc lexicon in SES than in NWS, and this is indeed what we find. I demonstrate this difference below by quantifying the degree to which basic vocabulary items in NWS and SES languages reflect reconstructable pOc etyma. I also examine the diversity of NWS lexicons, although there is no simple way to quantify this.

To determine the relative lexical conservatism of NWS and SES languages I calculated retention rates relative to pOc for most NWS languages, for a sample of SES languages (Bugotu, Gela, Tolo, Lau, Kwaio, Kwara’ae, To’aba’ita, Sa’a and Santa Ana), and, for comparison’s sake, for a few Oceanic languages outside the Solomons (Gedaged, Motu, Vitu, Tigak, Tabar, Lihir, Kandas, Mota and Bauan Fijian).

The procedure for calculating retention rates is modelled on that used by Blust (1981). He reconstructs Proto Malayo-Polynesian etyma for a modified version of the Swadesh 200-meaning list, calculating the percentage of these etyma reflected in each of the 55 languages in his database. Since all the 40 languages in my database are Oceanic, I instead used a baseline list of pOc etyma reflecting recent research, mainly Ross, Pawley and Osmond (1998, 2003) and, for free pronouns, Lynch, Ross and Crowley (2002:Ch.4). Following Blust (2000), I allowed more than one etymon per meaning where there is no discernible difference in meaning between reconstructed etyma. I modified Blust’s meaning list in various small ways and ran a trial with the resulting list of 199 meanings.

13 This inferred difference in social relations between Old Oceanic and MM speakers receives some support from Wickler’s (2001:241) interpretation of the archaeological sequence on Nehan and Buka islands in the North Bougainville area.

14 The list can be found at the Austronesian Basic Vocabulary Database website (http://language.psy.auckland.ac.nz/austronesian/). The website’s authors have added numerals from 6 to 10 and ‘a hundred’. I removed ‘this’, ‘that’, ‘and’, ‘no’, ‘if’, ‘other’ and ‘all’, as many Oceanic languages have more than one word corresponding to each and there is no obvious way of standardising one’s choice across languages. In keeping with Oceanic lexical organisation ‘we’ was replaced by ‘we (inclusive)’ and ‘we (exclusive)’ and ‘wife’ by ‘spouse’, and ‘salt’ was excluded as it is often not conceptually separable from ‘sea (water)’.

13
14
The trial revealed a number of weaknesses in the data set that a more constrained version of the meaning list could avoid. Space precludes a detailed account of these weaknesses, but they entailed ambiguity, polysemy and conceptual mismatch, and, as a result of these characteristics, there was either a plethora of reconstructions for a given item or insufficient or incorrect attestation of the required meanings in the wordlists. Meanings with these characteristics tend to display low retention in the database, but this must often reflect the fact that the words collected in different languages actually have different meanings, distorting retention percentages. In the light of these weaknesses, the list of 199 meanings was reduced to the list of 106 meanings shown in the Appendix and the analysis was repeated.

Table 1: Summary of retentions from Proto Oceanic for 106 meanings in 40 languages

<table>
<thead>
<tr>
<th></th>
<th>No. of meanings</th>
<th>No. of entries</th>
<th>Retention rate (%)</th>
<th>Reflexes of pOc items (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole database</td>
<td>102.4</td>
<td>112.7</td>
<td>46.8</td>
<td>49.7</td>
</tr>
<tr>
<td>North New Guinea</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gedaged</td>
<td>104</td>
<td>133</td>
<td>46.6</td>
<td>52.6</td>
</tr>
<tr>
<td>Central Papuan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motu</td>
<td>106</td>
<td>107</td>
<td>59.8</td>
<td>60.7</td>
</tr>
<tr>
<td>Bali-Vitu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitu</td>
<td>106</td>
<td>114</td>
<td>50.9</td>
<td>54.4</td>
</tr>
<tr>
<td>New Ireland</td>
<td>104.0</td>
<td>110.5</td>
<td>53.5</td>
<td>58.0</td>
</tr>
<tr>
<td>NW Solomonic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Bougainville</td>
<td>105.0</td>
<td>110.4</td>
<td>37.9</td>
<td>40.6</td>
</tr>
<tr>
<td>Banoni</td>
<td>106</td>
<td>132</td>
<td>35.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Mono-Torau</td>
<td>101.7</td>
<td>106.3</td>
<td>43.9</td>
<td>47.0</td>
</tr>
<tr>
<td>Choiseul</td>
<td>105.6</td>
<td>110.6</td>
<td>26.8</td>
<td>30.0</td>
</tr>
<tr>
<td>New Georgia</td>
<td>105.3</td>
<td>107.8</td>
<td>44.1</td>
<td>45.7</td>
</tr>
<tr>
<td>Santa Isabel</td>
<td>104.0</td>
<td>112.8</td>
<td>34.6</td>
<td>35.8</td>
</tr>
<tr>
<td>SE Solomonic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vanuatu/Banks-Torres</td>
<td>106</td>
<td>122</td>
<td>63.9</td>
<td>66.4</td>
</tr>
<tr>
<td>Mota</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central Pacific/Fijian</td>
<td>106</td>
<td>108</td>
<td>67.6</td>
<td>71.3</td>
</tr>
</tbody>
</table>

Table 1 summarises the analysis of retentions based on the 106-meaning list. Language names are in italics. One list, Piva, was excluded because it covered only 73 of the 106 meanings. Roman labels (‘North New Guinea’ etc) refer to groups of languages named in Lynch, Ross and Crowley (2002:Ch.5). There are four numerical columns. The first shows the (average) number of meanings (out of a possible total of 106) represented in the database for each language or group. The second shows the number of entries, which is usually greater than the number of meanings because of the inclusion of alternative items.

---

15 Conceptual mismatch refers to cases where the English meaning elicits more than one Oceanic concept, or no Oceanic concept at all.
The third shows the percentage of entries which are retained from pOc: that is, they reflect a pOc etymon with (more or less) the same meaning as that etymon, as explained above. The fourth column shows the percentage of entries that reflect a pOc etymon, regardless of whether the pOc meaning is retained. This is ignored in the remaining discussion.

The difference in retention rates between NWS and SES is dramatically clear from Table 1. The average retention rate for NWS languages is 36.6 percent, for SES 62.9 percent. None of the sample languages outside the Solomons displays a retention rate anywhere near as low as NWS. The lowest is the papuanised language Gedaged, at 46.6 percent (Motu, also papuanised, lies at 59.8 percent). As Table 2 shows, the highest retention rates occur in SES (Gela at 73.2 percent, Tolo at 70.4 percent) and in the sample languages from Remote Oceania (Bauan Fijian at 67.66 percent).

Table 2: Analysis of pOc retentions for 106 meanings in New Ireland and Solomons languages

<table>
<thead>
<tr>
<th>Language</th>
<th>No. of meanings</th>
<th>No. of entries</th>
<th>Retention rate (%)</th>
<th>Reflexes of pOc items (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Ireland</td>
<td>104.0</td>
<td>110.5</td>
<td>53.5</td>
<td>58.0</td>
</tr>
<tr>
<td>Tigak</td>
<td>101</td>
<td>110</td>
<td>51.8</td>
<td>56.4</td>
</tr>
<tr>
<td>Tabar</td>
<td>106</td>
<td>108</td>
<td>55.6</td>
<td>61.1</td>
</tr>
<tr>
<td>Lihir</td>
<td>106</td>
<td>110</td>
<td>58.2</td>
<td>62.7</td>
</tr>
<tr>
<td>Kandas</td>
<td>103</td>
<td>114</td>
<td>48.2</td>
<td>51.8</td>
</tr>
<tr>
<td>NW Solomonic</td>
<td>104.6</td>
<td>110.8</td>
<td>36.6</td>
<td>39.0</td>
</tr>
<tr>
<td>N Bougainville</td>
<td>105.0</td>
<td>110.4</td>
<td>37.9</td>
<td>40.6</td>
</tr>
<tr>
<td>Nehan</td>
<td>106</td>
<td>107</td>
<td>35.5</td>
<td>38.3</td>
</tr>
<tr>
<td>Solos</td>
<td>102</td>
<td>108</td>
<td>36.1</td>
<td>38.9</td>
</tr>
<tr>
<td>Haku</td>
<td>106</td>
<td>110</td>
<td>40.0</td>
<td>43.6</td>
</tr>
<tr>
<td>Teop</td>
<td>106</td>
<td>121</td>
<td>30.6</td>
<td>32.2</td>
</tr>
<tr>
<td>Taiof</td>
<td>105</td>
<td>106</td>
<td>47.2</td>
<td>50.0</td>
</tr>
<tr>
<td>Banoni</td>
<td>106</td>
<td>132</td>
<td>35.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Mono-Torau</td>
<td>101.7</td>
<td>106.3</td>
<td>43.9</td>
<td>47.0</td>
</tr>
<tr>
<td>Uruuva</td>
<td>95</td>
<td>97</td>
<td>44.3</td>
<td>47.4</td>
</tr>
<tr>
<td>Torau</td>
<td>104</td>
<td>111</td>
<td>44.1</td>
<td>47.7</td>
</tr>
<tr>
<td>Mono</td>
<td>106</td>
<td>111</td>
<td>43.2</td>
<td>45.9</td>
</tr>
<tr>
<td>Choiseul</td>
<td>105.6</td>
<td>110.6</td>
<td>26.8</td>
<td>30.0</td>
</tr>
<tr>
<td>Varisi</td>
<td>106</td>
<td>111</td>
<td>28.8</td>
<td>33.3</td>
</tr>
<tr>
<td>Vaghua</td>
<td>105</td>
<td>110</td>
<td>30.9</td>
<td>34.5</td>
</tr>
<tr>
<td>Ririo</td>
<td>106</td>
<td>110</td>
<td>26.4</td>
<td>30.0</td>
</tr>
<tr>
<td>Babatana</td>
<td>106</td>
<td>114</td>
<td>24.6</td>
<td>26.3</td>
</tr>
<tr>
<td>Sisingga</td>
<td>105</td>
<td>108</td>
<td>23.1</td>
<td>25.9</td>
</tr>
<tr>
<td>New Georgia</td>
<td>105.3</td>
<td>107.8</td>
<td>44.1</td>
<td>45.7</td>
</tr>
<tr>
<td>Simbo</td>
<td>103</td>
<td>108</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>Roviana</td>
<td>106</td>
<td>108</td>
<td>49.1</td>
<td>49.1</td>
</tr>
<tr>
<td>Hoava</td>
<td>106</td>
<td>107</td>
<td>37.4</td>
<td>39.3</td>
</tr>
<tr>
<td>Marovo</td>
<td>106</td>
<td>108</td>
<td>39.8</td>
<td>44.4</td>
</tr>
</tbody>
</table>
Table 2 shows the analysis language by language for New Ireland, NWS and SES. This reveals another difference between NWS and SES. Within the latter the highest retention rate is Gela at 73.2 percent, the lowest Bugotu at 52.2 percent (range = 21). Within NWS the highest is Simbo at 50 percent, the lowest Sisingga with 23.1 percent (range = 26.9). At first sight, it seems that the two groups have a similar profile, but that SES languages have retention rates around 23–29 percent higher than NWS. But a closer look reveals that this is not the whole story. Bugotu has a rather low retention rate by SES standards, the more so as its closest relative appears to be Gela, with a high 73.2 percent. The reason for the low rate in Bugotu is almost certainly that at Tataba village on the south-east tip of Isabel, Bugotu is spoken alongside Maringe. Tataba people are bilingual, and Bugotu has borrowed from Maringe, lowering its retention rate.¹⁶ A few such borrowings can be identified in the 106-meaning list:

(5) a. Bugotu kei- ‘tooth’; cf. Kok kei-, Kil khe’ñ-, Mge khe’ñ but Gela livo- < pOc *lipon

b. Bugotu ðehe ‘die’; cf. Kok Kil Mge lehe but Gela mate < pOc *mate

c. Bugotu sesëhu ‘grass’; cf. Sis sisú, Mvo tsetseu, Kia sesëhu but Gela yaojaboja

d. Bugotu kola- ‘liver’; cf. Bab Hoa Mvo Kia Kok kola-, but also, Bugotu ate-, Gela ate- < pOc *qate

<table>
<thead>
<tr>
<th>Language</th>
<th>No. of meanings</th>
<th>No. of entries</th>
<th>Retention rate (%)</th>
<th>Reflexes of pOc items (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isabel</td>
<td>104.0</td>
<td>112.8</td>
<td>34.6</td>
<td>35.8</td>
</tr>
<tr>
<td>Kia</td>
<td>106</td>
<td>112</td>
<td>32.1</td>
<td>32.1</td>
</tr>
<tr>
<td>Kokota</td>
<td>101</td>
<td>101</td>
<td>35.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Kilokaka</td>
<td>105</td>
<td>111</td>
<td>36.0</td>
<td>36.9</td>
</tr>
<tr>
<td>Maringe</td>
<td>104</td>
<td>127</td>
<td>34.6</td>
<td>35.4</td>
</tr>
<tr>
<td>SE Solomonic</td>
<td><strong>106.0</strong></td>
<td><strong>115.8</strong></td>
<td><strong>62.9</strong></td>
<td><strong>65.8</strong></td>
</tr>
<tr>
<td>Gelic-Guadalcanal</td>
<td>106.0</td>
<td>119.3</td>
<td>65.3</td>
<td>67.2</td>
</tr>
<tr>
<td>Bugotu</td>
<td>106</td>
<td>138</td>
<td>52.2</td>
<td>53.6</td>
</tr>
<tr>
<td>Gela</td>
<td>106</td>
<td>112</td>
<td>73.2</td>
<td>75.0</td>
</tr>
<tr>
<td>Tolo</td>
<td>106</td>
<td>108</td>
<td>70.4</td>
<td>73.1</td>
</tr>
<tr>
<td>Malaita-Makira</td>
<td>106.0</td>
<td>114.0</td>
<td>61.8</td>
<td>65.0</td>
</tr>
<tr>
<td>Lau</td>
<td>106</td>
<td>132</td>
<td>59.1</td>
<td>61.4</td>
</tr>
<tr>
<td>Kwaiio</td>
<td>106</td>
<td>108</td>
<td>53.7</td>
<td>57.4</td>
</tr>
<tr>
<td>Kwaraa’ae</td>
<td>106</td>
<td>114</td>
<td>63.2</td>
<td>66.7</td>
</tr>
<tr>
<td>Toabaita</td>
<td>106</td>
<td>108</td>
<td>61.1</td>
<td>64.8</td>
</tr>
<tr>
<td>Ña’ a</td>
<td>106</td>
<td>116</td>
<td>65.5</td>
<td>69.0</td>
</tr>
<tr>
<td>Santa Ana</td>
<td>106</td>
<td>106</td>
<td>67.9</td>
<td>70.8</td>
</tr>
</tbody>
</table>

¹⁶ There are also Santa Isabel borrowings in Gela and Tolo, but apparently fewer than in Bugotu.
Table 3: Variation in NW Solomonic retention rates

<table>
<thead>
<tr>
<th></th>
<th>No. of languages</th>
<th>Mean retention rate (%)</th>
<th>Range of retention rates (%)</th>
<th>Extent of range</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Ireland</td>
<td>4</td>
<td>53.5</td>
<td>48.2 – 58.2</td>
<td>10.0</td>
</tr>
<tr>
<td>NW Solomonic</td>
<td>22</td>
<td>36.6</td>
<td>23.1 – 50.0</td>
<td>26.9</td>
</tr>
<tr>
<td>New Georgia</td>
<td>4</td>
<td>44.1</td>
<td>37.4 – 50.0</td>
<td>12.6</td>
</tr>
<tr>
<td>Mono-Torau</td>
<td>3</td>
<td>43.9</td>
<td>43.2 – 44.3</td>
<td>1.1</td>
</tr>
<tr>
<td>N Bougainville</td>
<td>5</td>
<td>37.9</td>
<td>30.6 – 47.2</td>
<td>16.6</td>
</tr>
<tr>
<td>Banoni</td>
<td>1</td>
<td>35.6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Santa Isabel</td>
<td>4</td>
<td>34.6</td>
<td>32.1 – 36.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Choiseul</td>
<td>5</td>
<td>26.8</td>
<td>23.1 – 30.9</td>
<td>7.8</td>
</tr>
<tr>
<td>SE Solomonic</td>
<td>9</td>
<td>62.9</td>
<td>52.2 – 73.2</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Table 3 summarises from Table 2 the mean retention rates and ranges of retention rates for NWS languages (New Ireland and SES are shown for comparison). NWS groups are arranged in rank order of mean retention rates. Certain points emerge:

6. a. The rank order of mean retention rates bears no special relationship to geographical locations. Santa Isabel and Choiseul, with low retention rates, are in the east, but so is New Georgia, with a higher rate.

b. The highest retention rate in New Georgia is 50 percent, in North Bougainville 47.2 percent, suggesting that pNWS had a retention rate above 50 percent, i.e. a little higher than Kandas in Southern New Ireland with 48.2 percent (Kandas is the closest relative of pNWS included in the database).

The greater variation among retention rates of NWS subgroups suggests that the subgroups of NWS have more varied local histories than the two major subgroups of SES. This is supported by the observation in (6a), which implies that the differences between NWS subgroups reflect their histories more or less in situ. The observation in (6b) that pNWS had a retention rate above 50 percent, reflected in the protolanguages of the New Georgia and North Bougainville subgroups, means that losses of pOc reflexes leading to lower retention rates within these subgroups must have occurred independently within each subgroup. The Choiseul and Santa Isabel subgroups, however, display much less internal lexical variation, and suffered a reduction in their retention rates early in their separate histories. Proto Choiseul seems to have had a retention rate of, say, 33 percent, Proto Santa Isabel of around 38 percent.

The ranges of retention rates in North Bougainville and New Georgia suggest a chequered history even within these groups. Taiof displays the highest retention rate within North Bougainville, explained by Lincoln’s (1976a:422) observation that Taiof has been isolated from other North Bougainville languages — and has evidently had less contact with other languages than they have.
7 Shared lexical innovations in Northwest Solomonic groups

7.1 Distribution

I have attempted to reconstruct innovative lexical items for the 199-meaning list in the various Northwest Solomonic groups in order to gain an impression of how lexical innovations are distributed among them.

Table 4: Innovative lexical items in NW Solomonic groups in the 199-meaning list

<table>
<thead>
<tr>
<th>Languages</th>
<th>No. of languages</th>
<th>Mean retention rate from pOc (%)</th>
<th>No. of exclusively shared innovations</th>
<th>No. of innovations</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Georgia</td>
<td>4</td>
<td>44.1</td>
<td>38</td>
<td>77</td>
</tr>
<tr>
<td>Mono-Torau</td>
<td>3</td>
<td>43.9</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>N Bougainville</td>
<td>5</td>
<td>37.9</td>
<td>28</td>
<td>64</td>
</tr>
<tr>
<td>Banoni-Piva</td>
<td>2</td>
<td>35.6</td>
<td>12</td>
<td>37</td>
</tr>
<tr>
<td>Santa Isabel</td>
<td>4</td>
<td>34.6</td>
<td>59</td>
<td>92</td>
</tr>
<tr>
<td>Choiseul</td>
<td>5</td>
<td>26.8</td>
<td>41</td>
<td>91</td>
</tr>
<tr>
<td>W Choiseul</td>
<td>2–3</td>
<td>29.9</td>
<td>31</td>
<td>…</td>
</tr>
<tr>
<td>E Choiseul</td>
<td>2–3</td>
<td>23.9</td>
<td>24</td>
<td>…</td>
</tr>
</tbody>
</table>

The second and third columns of Table 4 are repeated from Tables 2 and 3. The fourth column shows the number of exclusively shared innovative lexical items in each group, supporting the claim that each is indeed a distinct subgroup. The fifth column shows the number of shared innovative lexical items in each group, including items reflected in more than one group, giving some sense of the degree of lexical innovation in NWS languages. ‘Innovative’ here means ‘not identifiable as Oceanic’, except in the case of the *-r-initial pronouns mentioned in §3. There is some expected inverse correlation of lexical innovations with retention rates from pOc. The Banoni-Piva figures are depressed in relation to figures for other groups because the Piva list covers only 87 meanings. Santa Isabel and Choiseul display 59 and 41 exclusively shared lexical innovations respectively (92 and 91 if we include innovations shared across group boundaries), which correlate with their low retention rates of 34.6 percent and 26.8 percent.

There is a complication in the Choiseul figures, as it became clear during the analysis that Choiseul falls lexically into two areas, West (Vagua and Varisi) and East (Babatana, Sisingga), which overlap in Ririo, implying that there was once a dialect chain stretching the length of the island. If we infer, as I do below, that lexical variation is largely due to lexical borrowings from different Papuan sources, then we must conclude that in earlier times there were two rather different Papuan languages on Choiseul. In addition to the 91 innovations which include West and East Choiseul, then, Table 4 shows 31 West Choiseul and 24 East Choiseul innovations.

17 There is a similar division in NGc, but my database contained only one WNGc list (Simbo), so I have not analysed the West/East difference here.
The differences between the figures in columns 4 and 5 of Table 4 indicate that a fair quantity of innovative items is shared between groups. This suggests the possibility that larger historic groupings of NW Solomonic languages might be identifiable on the basis of exclusively shared lexical innovations, and the relevant figures are presented in Tables 5 and 6. The first column of figures in Table 5 shows the number of exclusively shared lexical innovations in each pair of languages, and the remaining columns show the number in each trio. If there were larger historic groupings, then we would expect Table 5 to display rather larger numbers of exclusively shared lexical innovations than it does. In comparison with the numbers of innovations defining the six groups in Table 4, the figures in Table 5 are small, and suggest that each group shares a few innovations with its neighbours, as a result either of borrowing or of the differentiation of the NWS groups out of an earlier dialect network descended from pNWS.

A possible exception to this generalisation are the Choiseul, New Georgia and Santa Isabel groups. Choiseul and New Georgia share 13 innovations, New Georgia and Santa Isabel 10, and the three groups together 5. These innovations tell us that there is some kind of relationship between these three groups (with Mono-Torau sitting on the periphery) and that it is somewhat stronger than any other relationships apart from those within the six groups. But they do not tell us what kind of relationship it is. The shared innovations may indicate that the three groups have an exclusively shared common ancestor, but they may also reflect borrowing or an earlier relationship within an early NWS dialect network.

### Table 5: Innovative lexical items exclusively shared by two or three NW Solomonic groups (199-meaning list)

<table>
<thead>
<tr>
<th></th>
<th>–</th>
<th>MT</th>
<th>Ch</th>
<th>NGe</th>
<th>Is</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBv and BP</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NBv and MT</td>
<td>9</td>
<td>–</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>NBv and Ch</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NBv and NGe</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>NBv and Is</td>
<td>1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>BP and MT</td>
<td>5</td>
<td>–</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BP and Ch</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>BP and NGe</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>BP and Is</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MT and Ch</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MT and NGe</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>MT and Is</td>
<td>5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ch and NGe</td>
<td>13</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>5</td>
</tr>
<tr>
<td>Ch and Is</td>
<td>10</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NGe and Is</td>
<td>4</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 6: Innovative lexical items exclusively shared by four or more NW Solomonic groups (199-meaning list)

<table>
<thead>
<tr>
<th>Region</th>
<th>Count</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBv/BP/TM/Ch</td>
<td>2</td>
<td>*-r-o[e] ‘you SG’, *pisa ‘three’</td>
</tr>
<tr>
<td>NBv/BP/NGe/Is</td>
<td>1</td>
<td>*kolomo ‘water’</td>
</tr>
<tr>
<td>NBv/Ch/NGe/Is</td>
<td>1</td>
<td>*kapwa[ta] ‘skin’</td>
</tr>
<tr>
<td>MT/Ch/NGe/Is</td>
<td>2</td>
<td>*siko ‘steal’, *kavere ‘spider’</td>
</tr>
<tr>
<td>NBv/BP/Ch/Is</td>
<td>1</td>
<td>*siqop¥a ‘intestines’</td>
</tr>
<tr>
<td>NWS</td>
<td>1</td>
<td>*-r-[i]au ‘I’</td>
</tr>
</tbody>
</table>

Of the sparse innovations noted in Table 6, only one, *-r-[i]au ‘I’, is reflected in all six groups and is thus unambiguously reconstructable to pNWS. It seems very likely that the others, however, are also of pNWS antiquity but have been lost in one or two of the six groups in the course of their diverse histories. The same may also be true of some of the items for which reflexes have been found in only three groups.

These findings have consequences for reconstruction. To all intents and purposes NWS consists of six coordinate subgroups which are probably the result of a quite rapid spread of pNWS speakers through the smaller offshore islands and along the coastal strips of Bougainville, Choiseul, New Georgia and Santa Isabel. The resulting network of dialects diversified into today’s subgroups as NWS speakers interacted with speakers of a variety of Papuan languages. This process entailed the replacement of Oceanic etyma by Papuan loans at varying rates, as noted in §6. In these circumstances it would be otiose to insist that an etymon be reflected right across NWS in order to reconstruct in pNWS. Instead, I assume that any item that is reflected in Bougainville (in NBv, BP or MT) and in the northwest Solomons (in Ch, NGe or Is) is of pNWS antiquity. Because one may reasonably infer that pNWS was spoken on Buka, I also assume that any item reflected in NBv and at least one of BP and MT may also be reconstructed to pNWS. These criteria are loose for convenience. The looseness does not affect the arguments of this paper, but it does entail that ‘pNWS’ means ‘early NWS’ rather than exactly denoting the protolanguage.

7.2 Origins

Finding the origins of innovatory lexical items in NWS languages is no simple task. Given the two-wave proposal (§5), one might expect to find two kinds of origin: 18 borrowings from Papuan languages and borrowings from Old Oceanic languages. In addition, it is very likely that some unsourced items have an as yet unrecognised MM origin.

7.3 Lookalikes

Associated with the question of origins is the phenomenon of ‘lookalikes’. Lookalikes are lexical items in different NWS languages which are similar in meaning and form but which do not display regular sound correspondences. This means that they cannot be the direct result of shared inheritance, but are presumably the outcome of borrowing at some point.

18 There are also a few words reflecting etyma in MM interstage languages of a higher order than pNWS, for example, eMM *qase(n), *qasen-i- ‘count’ (Vitu yadenti-, Tabar ase, Neh ah, Teo ahe, Mon kala, Mvo ase, Kil a:ahe). Such items are not included in the figures in Tables 4 and 5.
Pairs of lookalikes fall into two categories. The first category consists of a lexical item directly inherited from pOc (or from an interstage later than pOc but earlier than pNWS) paired with an item that appears to be descended from the same pOc etymon but via borrowing. For example, the items in (7a) are fairly regular reflexes of pOc *qalipan ‘centipede’. The minor irregularities they display are not likely to be due to borrowing.\(^{19}\)

The items in (7b), however, reflect a putative pNWS *kalivaga, where pOc *q- is irregularly reflected as *k- and pOc *-n as *-g-. One may reasonably infer that this was borrowed into an early NWS dialect from another Oceanic language, probably an Old Oceanic language, in which pOc *q-was reflected as *k-.

(7) a. pOc *qalipan ‘centipede’ > pNWS *aliva
   \[\text{Neh liha-}, \text{Tai aifan}, \text{MT Uru rivana WNGe Lun li-liva-}, \text{Ndu livanya, Is Bla n-alifa, Gho Mge n-alhija}\]
   b. pNWS *kalivaga > NBv Tai zanevaga, ECh Rir Bab Sis kaligava (metathesis)

The probability of a chance similarity between (7b) and (7a) is very low indeed, given that the pOc etymon had four consonants and three vowels, each needing to be matched by chance. The question with lookalikes, of course, is the point at which chance becomes a major factor. One could establish criteria based on the concepts in Nichols (1996:50–54), but this would entail discussion beyond the scope of this paper. Intuitively, however, it is obvious that (8a) and (8b), with fewer resemblant syllables, are more likely to be outcomes of chance resemblances — but may also indirectly reflect *qalipan.

(8) a. NGe Mvo lipata ‘centipede’
   b. pNWS *kali ‘centipede’ > NBv Teo kare, MT Mon ale-le, WCh Var kali-kali

Lookalikes like the pair in (8) are based on a set of inferences similar to those made by Biggs (1965) about the history of Rotuman, where one inherited and two borrowed layers of vocabulary are distinguished, or by Ross (1996) about Yapese, where five contributing sources, one inherited and four borrowed, are identified. The difference, however, is that the layers of vocabulary in NWS are of greater antiquity, and the source language(s) of one of the layers cannot be reconstructed.

There is strong evidence to infer that NWS languages have borrowed from Papuan neighbours, and it is possible that an item like (7b) was borrowed from a Papuan language which had borrowed it from an Old Oceanic source.

The second category of lookalikes consists of pairs neither of whose members can be sourced to pOc. This may mean one of several things (apart from chance resemblance). The items may be directly and indirectly descended from an as yet unreconstructed pOc etymon, or they may reflect different borrowings of a single Papuan item, or they may reflect borrowings of cognate items from two different Papuan languages. For example, pNBv *ma(l,r)oto in (9a) and pNWS *manoqa in (9b), both ‘ten’, look as if there is some commonality in their history: the first syllable of both is *ma-, the second is an apical plus *-o-. But the apicals don’t correspond and the final syllables don’t match, suggesting that the commonality is mediated by borrowing, perhaps of items for ‘ten’ from different but related Papuan languages. The fact that the pOc term for ‘ten’ in (9c) also survived — and

\(^{19}\) pNWS *-ŋ- is an idiosyncratic reflex of pOc *-n (Ross 1988:223); Nehan, Haku, Uruava, Lungga and Nduke have lost initial *qa- (cf. pOc *qapaRa ‘shoulder’ > pNWS *para > Hak hala-hala, Tor ara, Mon hala, Lun Ndu Rov vara, Lag fara); and Taiof loss of *-l- is unusual.
its reflexes coexist with those of *ma(l,r)oto in the NBv group — reinforces the possibility that borrowings were often very localised.

(9) a. pNBv *ma(l,r)oto ‘ten’ > Sol manot, Pet malot, Hak maloto, Sel malto
   b. pNWS *manoya ‘ten’ > BP Ban mano ya, Ch Var mano ya, Rir manua, Bab mano, Sis mano, NGe Sim mano ya, Lun Ndu mano ya puta, Rov Hoa mano yeputa
   c. pOc *saŋapulu ‘ten’ > PNS *saŋavulu > NBv Neh haŋaulu, Tai safunu, Teo savun, MT Tor saumu, Mon lahulu, Uru avuru, WCh Vag ŋəvəl

7.4 Borrowings from Papuan languages

I have sourced no Papuan loanwords to my satisfaction.

Three groups of present-day Papuan languages are relevant to the search for sources. Much of Bougainville is occupied by the North and South Bougainville families. Lincoln (1976b) investigated the claim that Piva basic vocabulary includes loans from neighbouring Rotokas (N Bougainville). He found no evidence for them. The South Bougainville family falls into two subfamilies, Nasioi and Buin, represented respectively by dictionaries of Nasioi (Hurd and Hurd 1974) and Buin (alias Telei; Laycock 2003). The most likely candidate for borrowings from the Nasioi subfamily is Banoni, and from the Buin subfamily, Mono, but I have found no evidence of such loans.

The third Papuan group is the Central Solomons family, consisting of four geographically scattered languages: Bilua, Baniata, Lavukaleve and Savosavo. The first two are located in the New Georgia group and are plausible sources for borrowings in New Georgia languages. Bilua is also a plausible borrowing source for Choiseul languages. The wordlists in Tryon and Hackman (1983) show a number of borrowings involving each language, but in a number of cases the direction of borrowing is clearly from Oceanic to Papuan, and there is no unambiguous evidence of borrowing in the opposite direction.

The fact that there are so many unsourced NWS etyma is less puzzling than it may seem. A careful look at Todd’s (1975) wordlists for the Central Solomons family reveals that almost the only recognisable cognates among them are Oceanic loans. The Papuan languages of north-west island Melanesia have been in situ for such a long time that their basic vocabularies have diverged beyond recognition. This is true not only of the Central Solomons languages, but also of the North and South Bougainville families. The three families are apparently unrelated to each other, and relationships within each family seem rather distant, a reflection of their great time depth (Ross 2001; Dunn, Reesink and Terrill 2002). It can be readily inferred that before the arrival of Oceanic speakers in the region, there were far more Papuan languages in north-west island Melanesia than there are today, and that they already reflected a great degree of diversity. Dunn et al. (2005) have suggested that their diversity dates back more than 10,000 years. If this is so, then it is eminently likely that the present-day Papuan languages referred to in this section bore no recognisable relationship to the now lost languages that contributed vocabulary to early interstages of NWS and that many NWS lexical items will thus remain unsourced.

---

20 Sometimes called the West and East Bougainville families.
7.5 Borrowings from Old Oceanic languages

As noted in §7.3, there are a number of lookalikes where one member of the pair is directly inherited from pOc and the other is an indirectly inherited reflex of the same etymon, apparently borrowed from another (Old) Oceanic language. Since these lookalikes are evidence for the hypothesis that there was Old Oceanic settlement in the NWS region before the MM settlement reflected by NWS itself, the most clearly attested of them are set out below. One, (8), was given in §7.3.

Whereas the forms in (10a) reflect normal pNWS lenition of pOc *-k- as *-ɣ- and loss of pOc *-w-, the forms in (10b) reflect an Oceanic form in which pOc *-k- is unlenited and *-iw- is preserved as *-u-.

(10) a. pOc *bakiwa ‘shark’ > pNWS *baɣea > Bab Sis bəza, Gha Lun Sim bəya, Kil Kok bæe-su, Mge baɬe-su
b. pNWS *bakuai ‘shark’ > NBv Neh bakue, Sol bake, Hak baki, Sel buei, Teo baku-baku, BP Ban bakuo, MT Tor vavoi, Uru baku-baku, Mon baɬoi, WCh Vag bakuai, Var bakuai, Rir boɬoei

The forms in (11a) are regular reflexes of pOc *pituqun, whereas those in (11b) reflect an Oceanic form with reduplication, unlenited *p-, loss of *-q- and *-n. In (11c) pECh *-putu seems to be a separate Old Oceanic reflex. In (11d) pNGe *pi(no)-pino raises the classic lookalike problem: is it an Old Oceanic reflex (with loss of *-t- and retention of *-n-?) or does it resemble pNWS *pi(to)-pito by chance?

(11) a. pOc *pituqun ‘star’ > pNWS *vitu(ɣu)nu > MT Uru vesunu, Is Gho Mge natʰunu
b. pNWS *pi(to)-pito ‘star’ > NBv Neh pito-pit, Sol bi-pit, Hak pito-pito, BP Ban Piv pi-pito, MT Mon vito-vito, Tor vi-vito
c. pECh *sisiri-putu ‘star’ > Bab Sis sisiri-putu
d. pNGe *pi(no)-pino ‘star’ > Ndu Ugh pi-pino, Rov Kus pinopino

The forms in (12a) are regular reflexes of pOc *boRok, whereas those in (12b) are open to two analyses: they reflect either (i) loss of pOc *-R- and unlenited *-k, or (ii) pOc *-R-as *-k- and loss of pOc *-k. Papuan languages have also borrowed this term from Oceanic: Bilua bolo appears to be a borrowing from WNGe boroyo, Baniata bo from ENGe boko.

(12) a. pOc *boRok ‘pig’ > pNWS *boroyo > BP Ban boroyo, WNGe Lun Ndu boroyo
b. pNWS *boko > NBv Tai vo, MT Tor bo, Mon boɬo, Ch Vag Var boko, Rir boʔ, Sis Bab boko, ENGe Rov Hoa book

The Santa Isabel reflexes in (13a) appear to be directly inherited, as do the Piva (BP) and Sim (WNGe) reflexes of the same form with prefixed *[ma]- in (13b). However, pCh *madaka- in (13c) appears to be an indirectly inherited reflex of this prefixed form, with the same diachronic phonological ambiguity as pNWS *boko ‘pig’ in (12b). If pCh

---

21 Reflexes of *bakuai ‘shark’ are also found in New Ireland (Madak bokiǔ ‘dugong’, Tolai boko, Konomala bakui, Siar baki) alongside regular reflexes of *bakiwa, suggesting that there was also a dialect or dialects there which predated MM.
22 Here -z- appears to reflect *-y-: *bəya > *baɣea > *baya.
23 Aspirated -ɬh- reflects pIs *-f- (< pOc *pVt-).
*madaka- is indeed indirectly inherited (i.e. an Old Oceanic loan), then the possibility exists that *[ma]madara(ya)- in (13b) and even *dara(ya)- in (13a) are also indirectly inherited.

(13) a. pOc *draRaq ‘blood’ > pNWS (?) *dara(ya)- > pIs Mge Kok Kil dadara
   b. pNWS (?) *[ma]ma-dara(ya)- > BP Piv ma-rana-, WNGe Sim mama-dara
   c. pCh *madaka- > Vag Bab Sis madaka-, Rir madak

I have included this example not because I am especially optimistic about the suggested analysis above but because it gives some flavour of the analytical problems that NWS lexicon confronts us with. Also found are Rov ehara-, Mvo Van juka-, Hoa Kus mazuka- and Kia busaka-, all ‘blood’. The (Papuan) Bilua term is dara-, surely borrowed from an Oceanic language, Old or Western, but there is no evidence of the form in the usual Oceanic sources of Bilua borrowings.

Returning to more straightforward examples, (14a) contains regular reflexes, whilst (14b) and (14c) appear to have been borrowed from Old Oceanic languages at different points in NWS history. The former has voiced *b- and retains unlenited *-k-, and the latter has voiceless *p- and loses *-k-.

(14) a. pOc *bekas ‘defecate’ > pNWS *beɣasa > NBv Neh beh,24 BP Ban beɣasa, WCh Vag bɪɣa, Var beɣa
   b. pNBv *beka > Teo bekëka
   c. PWS *pea > MT Mon pea, ECh Rir Bab Sis pia, NGe Gha Lun Ndu Rov Van pea

In (15) there appear to be two pNWS forms. The first reflects pOc *wakaR regularly with loss of *w-. The second reflects *w- as *b- and was apparently an Old Oceanic form.

(15) a. pOc *wakaR ‘root’ > pNWS *aɣara- > Teo ana, Sel ara, Tor agarə-, Uru agarə-, NGe Ndu ayara, Hoa ayoro, Is Kia Kok Kil zagra
   b. pNWS *baɣara- NBv Tai vora-, BP Ban baɣura-, Piv bagara-, WNGe Sim Lun bayerə, Is Lag bakla

The examples above stand out because the pOc forms have three consonants and 2–3 vowels and doublet reflexes, as well as reflexes from across the NWS area, i.e. there is adequate evidence that their doublets reflect borrowings from Old Oceanic forms. There are a number of other forms that do not satisfy these criteria, and varying degrees of doubt must subsist as to whether they are reflexes of Old Oceanic forms, chance resemblances, or even irregular reflexes of MM forms. Some of these forms have indeed been analysed in the past as irregular reflexes of MM forms.

In Ross (1988:224) I remarked that pOc/pNWS *s became Proto West Choiseul25 *j in some lexical items. My examples were (16) and (17). Notably pOc *siku is apparently not reflected elsewhere in NWS languages, increasing the likelihood that pWCh *jiku was a local borrowing. Both items reflect unlenited pOc *k rather than the usual inherited *ɣ. I return to reflexes of pOc *kusupe below.

24 Neh -h reflects *-s.
(16) pOc *siku ‘elbow’ > pWCh *jiku > Vag zə-zəkə, Varisi zi-ziku
(17) pOc *kusupe ‘rat’ > pWCh *kju > Vag kəj, Var kuzu, Rir kuj

There are numerous other lookalikes in Choiseul languages, correlated with the fact that they have the lowest retention rates in NWS, but it is often hard to be sure that they are not chance resemblances. For example, Tryon and Hackman (1983:61) note pCh *pade ‘house’ as an irregular reflex of pOc *pale. It is likely that they are right, and that this is an Old Oceanic loan, but the possibility that it is a chance resemblance cannot be ruled out.

There are a number of apparent Old Oceanic borrowings limited to the Santa Isabel group. The doublet form in (18b) is reduplicated and displays loss of pOc *-t-.

(18) a. pOc *mataqut ‘fear’ > pWCh *matayu > NBv Pet matout, Hak matatu, NGe Sim matatu, Gha Lun Rov mataγut
b. pIs *mamayu > Ki mamaγu, Kil Mge hmaγu

The items in (19) and (20) are given by Tryon and Hackman (1983:61) as illustrations of the Santa Isabel reflex of pOc *w, but there is good evidence that pOc *w was lost in the Santa Isabel group, as in other NWS languages (Ross 1986:186–197). It thus seems likely that Santa Isabel items like these, in which pOc *w has merged with *p, represent Old Oceanic loans (this does not belie Tryon and Hackman’s basic point, that these items are uniquely shared innovations of Santa Isabel languages).

(19) a. pOc *ma-nawa ‘breathe’ > pNWS *ma-naa ‘heart’ > Is Gho ma-ñaα ‘heart’

(20) a. pOc *siwa ‘nine’ > pNWS *sia > NBv Neh lu-sio, Sol sie, Hak to-si, Tai sia, MT Tor sia, Mon u-ia, Uru ia, Ch Var ka-ia, Rir zia, Bab zia, Sis zia, NGe Sim Rov Hoa sia
b. pIs *n-heva > Koγ heva, Kil nhe ı̆va, Mge nhevai

The two lookalike sets below further exemplify the difficulty of unravelling NWS lexical history. pNWS *tuyur, the regular reflex of pOc *tuqur, is well attested in (21a). It is tempting to assume that pMT *tegese in (21b), pCh *deyere in (21c) and pIs *tetu in (21d) are all borrowings of Old Oceanic reflexes of pOc *tuqur, but again they may be chance resemblances.

(21) a. pOc *tuqur ‘stand’ > pNWS *tuyur > Neh tur, Sol tonon, Teo sun, Tai tutun, Ban tsuyono, Uru toru, Gha Lun Sim Ndu Rov turu
b. pMT *tegese > Mon Tor tegese
c. pCh *deyere > Var Vag deyere, Rir der, Sis dire, Bab dere
d. pIs *tetu > Kok Kil Kia

---

26 Tryon and Hackman (1983:60) point out that when an initial consonant is lost before *-i- in Choiseul, z-accretion occurs in all languages except Vaghua. Here, however, z- also occurs in Vaghua, indicating that this is not accretion.
The forms for ‘rat’ provide even greater difficulties. We know that pOc *kusupe ‘rat’ was reflected as eMM *kusuve, i.e. with an unlenited initial *k-, as reflected in the MM forms in (22) from languages outside NWS.27

(22)  pOc *kusupe ‘rat’ > eMM *kusuve > Bali-Vitu Vitu kuvude (*sl*s metathesis), Bali kuvuzeke (*sl*s metathesis), Willaumenz Bulu Bola kuruve, Nakanai kusuke, Tabar Lihir kues, Madak Lamasong Madak kisap, Barok kisu, S New Ireland Konomala kusi, Minigir kusua, Label Siar kusu, Bilur kue, Kandas kusu, Ramoaaina kaupa

The expected pNWS form would also have been *kusuve, but there are no forms with a reflex of *-v-, and so I reconstruct pNWS *kusu(v)equ. Proto North Bougainville *kiso in (23a) is one of several lexical items in which pOc *u became pNBv *i (Ross 1988:223), and is quite possibly directly inherited. Proto Santa Isabel *kusi in (23c) is also a plausible directly inherited reflex (cf. Konomala kusi in (22)), which has also been borrowed as Bugotu, Gela (SES) kuhu.

(23)  pOc *kusupe ‘rat’ > eMM *kusuve > pNWS *kusuv(e) >
    a.  pNBv *kiso > Neh kih, Sol kiso, Hak isu, Teo kuho, Tai kiso
    b.  pBP *kiso > Ban kiso
    c.  pIs *kusi > Mge na-khusi

Lookalikes are listed in (24). Proto Mono-Torau *kuake was perhaps an Old Oceanic borrowing, in which pOc *-s- has been lost and *-p- had been replaced by *-k- (cf. Nakanai kusuke in (22)). Proto New Georgia *karuje is not readily explicable, but pECh *ruji is plausibly a version of *karuje with first-syllable loss. Proto East Choiseul *kuju was discussed above. It is possible that pENGe *kutu was a borrowing of an Old Oceanic form cognate with the source of pECh *kuju. However, similar forms are found in New Ireland (Tabar kotu, Tangga kutu) and it is also possible that pENGe *kutu was cognate with them. This reminds us, incidentally, that New Ireland may also have had an Old Oceanic period, also reflected in lookalikes.28

(24)  a.  pMT *kuake > Mon Tor kuake, Uru kue
    b.  NGe *karuje > Sim karuje, Ndu Rov kurezu
    c.  pECh *ruji > Bab ruji, Sis roji
    d.  pWCh *kuju > Vag kaj, Var kuzu, Rir kuj (repeated from (17))
    e.  pENGe *kutu > Hoa Mvo cutu

The variation in terms for ‘rat’ is so great that I have also wondered whether they represent later borrowings after MM/NWS settlement. However, this would presuppose that Rattus exulans followed Oceanic speakers into the region. As Matthew Spriggs comments (pers. comm. 2007), ‘the route for Rattus exulans into the Pacific was clearly as stowaways or snack lunches on Lapita canoes … The other complication is of course the rich endemic rat fauna of the main Solomons compared to the Bismarck Archipelago …’ A more likely explanation of the terms above is that they represent parallel borrowings from Old Oceanic languages.

27 The pSES form, attested without change in West Guadalcanal and Talise was *yusuve with lenited *γ-
28 Compare Tigak kusia, Tiang kuse, Kara kuf, all ‘rat’. 

What conclusions can we draw from the sample of lookalikes in this section? Examples (7) and (10–15) are generally more convincing evidence for the erstwhile presence of Old Oceanic languages in the NWS region than (16–24). There is a methodological difficulty here. The probability of chance resemblances is less with regard to three- and four-syllable forms than it is with regard to two-syllable forms — but the majority of pOc roots, at least in more basic vocabulary, had only two syllables. Reinforcing the case for Old Oceanic requires a large number of examples, and these will only be forthcoming, if ever, when detailed dictionaries are available for a good sample of NWS languages. For the moment we can say simply that the evidence favours an Old Oceanic presence in the region, and that certain pNWS etyma appear to be Oceanic but not to be directly inherited from pOc. I have also suggested that there were later, local borrowings from Old Oceanic sources, particularly in Choiseul and Santa Isabel, but this suggestion in particular requires more research.

8 Conclusions

I have shown that retention rates in NWS languages are lower than elsewhere in Western Oceanic and considerably lower than in SES languages. It is clear from the differing retention rates of NWS groups, however, that the groups must have had rather different histories. This is confirmed by the relative absence of both morphosyntactic and lexical innovations defining larger groupings within NWS, and the resultant need to recognise six primary groups within NWS.

In what respect did the histories of these six groups differ? The answer would seem to be that their speakers have all been in contact with speakers of other languages and been bilingual in those languages at various periods in their histories. Differing retention rates reflect different degrees of contact or different social relationships with speakers of other languages. Frustratingly, we have virtually no clear evidence for loans from Papuan languages, but simply large numbers of etyma for which there seems to be no other explanation. We do have some evidence, however, for loans from Old Oceanic languages. One intriguing feature of the data and reconstructions in §7.5 is that different Oceanic reflexes occur cheek by jowl with one another (and with apparent Papuan loans) in languages of the same group. The only reasonable explanation I can offer for this is that even quite local contact histories varied. This seems clearly to have been true of west and east Choiseul and of west and east New Georgia, as well as of the three groups on Bougainville. Indeed, the thought with which I would like to finish this paper is simply that there is huge lexical diversity among NWS languages, despite their evident genealogical unity, and this must be due largely to differing contact histories. This claim will stand, I think, even if some of the detailed proposals here are replaced by better ones.
Appendix: Reduced meaning list, with percentage Proto Oceanic retentions for each meaning (ID numbers match those used in the Austronesian Basic Vocabulary Database)

<table>
<thead>
<tr>
<th></th>
<th>Meaning</th>
<th>Retention</th>
<th></th>
<th>Meaning</th>
<th>Retention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hand</td>
<td>56.1</td>
<td>56</td>
<td>child</td>
<td>27.3</td>
</tr>
<tr>
<td>2</td>
<td>left</td>
<td>48.7</td>
<td>59</td>
<td>mother</td>
<td>61.4</td>
</tr>
<tr>
<td>3</td>
<td>right</td>
<td>43.6</td>
<td>60</td>
<td>father</td>
<td>84.4</td>
</tr>
<tr>
<td>4</td>
<td>leg/foot</td>
<td>40.5</td>
<td>61</td>
<td>house</td>
<td>52.4</td>
</tr>
<tr>
<td>5</td>
<td>to walk</td>
<td>30.6</td>
<td>63</td>
<td>name</td>
<td>75.6</td>
</tr>
<tr>
<td>6</td>
<td>road/path</td>
<td>60.0</td>
<td>64</td>
<td>to say</td>
<td>6.3</td>
</tr>
<tr>
<td>7</td>
<td>to come</td>
<td>87.8</td>
<td>65</td>
<td>rope</td>
<td>39.6</td>
</tr>
<tr>
<td>11</td>
<td>dust</td>
<td>35.0</td>
<td>74</td>
<td>to kill</td>
<td>22.2</td>
</tr>
<tr>
<td>12</td>
<td>skin</td>
<td>35.7</td>
<td>75</td>
<td>to die</td>
<td>72.5</td>
</tr>
<tr>
<td>14</td>
<td>belly</td>
<td>33.3</td>
<td>76</td>
<td>to be alive</td>
<td>61.5</td>
</tr>
<tr>
<td>15</td>
<td>bone</td>
<td>45.0</td>
<td>78</td>
<td>to cut</td>
<td>15.5</td>
</tr>
<tr>
<td>16</td>
<td>intestines</td>
<td>16.7</td>
<td>79</td>
<td>stick/wood</td>
<td>58.1</td>
</tr>
<tr>
<td>17</td>
<td>liver</td>
<td>50.0</td>
<td>90</td>
<td>to dig</td>
<td>39.5</td>
</tr>
<tr>
<td>18</td>
<td>breast</td>
<td>86.7</td>
<td>95</td>
<td>to fall</td>
<td>29.8</td>
</tr>
<tr>
<td>22</td>
<td>to fear</td>
<td>27.9</td>
<td>96</td>
<td>dog</td>
<td>15.9</td>
</tr>
<tr>
<td>23</td>
<td>blood</td>
<td>50.0</td>
<td>97</td>
<td>bird</td>
<td>39.0</td>
</tr>
<tr>
<td>24</td>
<td>head</td>
<td>52.4</td>
<td>98</td>
<td>egg</td>
<td>30.0</td>
</tr>
<tr>
<td>27</td>
<td>nose</td>
<td>72.5</td>
<td>99</td>
<td>feather</td>
<td>34.9</td>
</tr>
<tr>
<td>28</td>
<td>to breathe</td>
<td>38.1</td>
<td>101</td>
<td>to fly</td>
<td>46.3</td>
</tr>
<tr>
<td>29</td>
<td>to sniff</td>
<td>29.5</td>
<td>102</td>
<td>rat</td>
<td>25.6</td>
</tr>
<tr>
<td>31</td>
<td>tooth</td>
<td>62.5</td>
<td>103</td>
<td>meat/flesh</td>
<td>47.6</td>
</tr>
<tr>
<td>32</td>
<td>tongue</td>
<td>87.5</td>
<td>105</td>
<td>tail</td>
<td>17.9</td>
</tr>
<tr>
<td>33</td>
<td>to laugh</td>
<td>5.0</td>
<td>106</td>
<td>snake</td>
<td>30.0</td>
</tr>
<tr>
<td>34</td>
<td>to cry</td>
<td>61.9</td>
<td>108</td>
<td>louse</td>
<td>82.5</td>
</tr>
<tr>
<td>35</td>
<td>to vomit</td>
<td>84.6</td>
<td>109</td>
<td>mosquito</td>
<td>23.3</td>
</tr>
<tr>
<td>36</td>
<td>to spit</td>
<td>68.3</td>
<td>110</td>
<td>spider</td>
<td>15.6</td>
</tr>
<tr>
<td>37</td>
<td>to eat</td>
<td>56.3</td>
<td>111</td>
<td>fish</td>
<td>67.5</td>
</tr>
<tr>
<td>40</td>
<td>to drink</td>
<td>32.6</td>
<td>113</td>
<td>branch</td>
<td>31.7</td>
</tr>
<tr>
<td>41</td>
<td>to bite</td>
<td>51.1</td>
<td>114</td>
<td>leaf</td>
<td>50.0</td>
</tr>
<tr>
<td>43</td>
<td>ear</td>
<td>75.6</td>
<td>115</td>
<td>root</td>
<td>46.8</td>
</tr>
<tr>
<td>44</td>
<td>to hear</td>
<td>78.0</td>
<td>117</td>
<td>fruit</td>
<td>59.5</td>
</tr>
<tr>
<td>45</td>
<td>eye</td>
<td>92.5</td>
<td>118</td>
<td>grass</td>
<td>12.5</td>
</tr>
<tr>
<td>48</td>
<td>to sleep</td>
<td>22.7</td>
<td>119</td>
<td>earth/soil</td>
<td>31.7</td>
</tr>
<tr>
<td>53</td>
<td>person</td>
<td>38.6</td>
<td>120</td>
<td>stone</td>
<td>48.8</td>
</tr>
<tr>
<td>54</td>
<td>man/male</td>
<td>52.4</td>
<td>121</td>
<td>sand</td>
<td>31.0</td>
</tr>
<tr>
<td>55</td>
<td>woman/female</td>
<td>27.5</td>
<td>122</td>
<td>water</td>
<td>29.3</td>
</tr>
</tbody>
</table>

References


—— 2007a, Recent research on the historical relationships of the Papuan languages, or, what does linguistics say about the prehistory of Melanesia? In Friedlaender, 2007, 36–59.


What can the Mon-Khmer lexical borrowings in Acehnese tell us?

PAUL SIDWELL

1 Introduction

It is a great pleasure to honour Professor Andy Pawley on the occasion of his retirement, given the tremendous service he has rendered the field of Linguistics, and colleagues and students generally, over many years. My own career has benefited crucially from the support that Andy has shown me over nearly a decade, and I happily dedicate this short paper to him.

To summarise as briefly as possible, since Niemann (1891) there has been an off-and-on discussion concerning how closely related Acehnese is to Chamic (as opposed to, say, Malayic), and in connection with this, to what extent the non-Austronesian characteristics of Acehnese can be explained by contact with Mon-Khmer (MK) languages, either independently, or reflecting descent from a contact-affected Proto Aceh-Chamic. The most recent and elaborate study is that of Thurgood (1999) (henceforth ‘Thurgood’). Thurgood offers a comprehensive historical-phonological framework within which the Acehnese and Chamic data can be assessed. His study hints strongly at an historical account which locates Proto Aceh-Chamic in Indo-China during the first millennium C.E., with the origin of Acehnese in an emigration from northern Champa in response to the expansion southward of Vietnam. However, Thurgood does not treat Acehnese lexical data as extensively as he does Mainland Chamic, leaving a gap that needs to be addressed.

The approach I am now taking is to attempt a more-or-less comprehensive account of the problematic Acehnese lexicon, on the assumption that distinct patterns among borrowings may identify donor languages, locate the historical Acehnese speakers geographically, and explain the origins of structural/typological features. With the recent (2006) appearance of Shorto’s A Mon-Khmer comparative dictionary, and the online lexical resources of the Mon-Khmer languages project of SEALANG (http://sealang.net/monkhmer/) it should now be feasible to attempt a preliminary etymological study.

As a first step I went through the exercise of extracting and consolidating the Acehnese lexicon of likely MK origin identified in the following sources:

1 Work which contributed to this paper was made possible by financial assistance from the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology (Leipzig), and the National Endowment for the Humanities (Washington). Any views, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.
Cowan (1948): contains a lexicon of 150 numbered Acehnese items with MK and other comparisons,

Shorto (1975): discusses the classification of Acehnese vis-à-vis Chamic and Malay with numerous lexical examples,

Collins (1975): unpublished PhD dissertation lists more than 1500 Acehnese words with comparisons from various languages of the region, and

Thurgood (1999): a reconstruction of Proto Chamic which treats Acehnese as a Chamic language.

Thurgood is not extensive in his citation of relevant Acehnese lexic — most of those he lists appear to be recompiled from Cowan (1948) without a large body of new additions — his index of about 270 Chamic etymologies of MK origin includes only 44 Acehnese members (a point made by Dyen in his 2001 review). The somewhat hard to obtain (if you are not handy to Berkeley, California) dissertation of Collins (1975) is superficially promising, offering lexical comparisons for 1500 Acehnese words, but subsequent analysis shows that hardly more than a hundred of these are meaningfully useful.2

From these sources I compiled a database of 145 Acehnese words of likely MK origin, and these are organised and listed below.

2 Lexicon and analysis

The lexical comparisons compiled for this study can be stratified as follows:

(1) 50 Aceh-Chamic isoglosses with MK, presumably most are borrowed from MK,

(2) 18 sets of apparently independent Acehnese and Chamic borrowing of the same or related MK roots,

(3) 63 Acehnese-MK isoglosses apparently without Chamic cognates, and

(4) 14 unique Acehnese-Nicobarese isoglosses,

Broadly speaking the gross numbers hint at an interesting pattern: there nearly 50 MK borrowings that belong to the reconstructable Proto Aceh-Chamic vocabulary, clearly indicating some period of linguistic unity. In addition to these, there are more than 90 MK loans, or nearly two thirds of the identified MK borrowing, which were apparently received after the separation from Chamic, indicating a considerable history of separate development while remaining in contact with MK languages.

2.1 Aceh-Chamic isoglosses

The Aceh-Chamic isoglosses tentatively identified as belonging to the Proto Aceh-Chamic period can be further stratified according to the distribution of cognates among MK languages, or the lack of cognates for that matter. The first group below share cognates which are widely distributed in MK, and lack phonological or semantic clues linking them to any particular branch or language. (Note: Proto Chamic (PC) forms are from Thurgood, Proto Mon-Khmer (PMK) from Shorto (2006), and glosses are not entered if they are the same as for the previous cognate form. Forms in bold are phonetic/phonological representations, and those in italics are orthographic representations or their transcriptions.)

2 Collins employed excessively lax semantic and phonological criteria in his comparisons.
<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>MK reflexes widely distributed geographically</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dit</strong> ‘few’</td>
<td>PC *dVt ‘small’</td>
<td>PMK 1016 *kift etc.</td>
</tr>
<tr>
<td><strong>gap</strong> ‘firm’</td>
<td>Cham kàp ‘enough, just right’</td>
<td>PMK 1240 *gap etc. ‘fit, fitting, sufficient’.</td>
</tr>
<tr>
<td><strong>gap</strong> ‘other, group’</td>
<td>PC *gap</td>
<td>PMK 1241 *gap, *gap ‘friend, to associate’.</td>
</tr>
<tr>
<td><strong>ja</strong> ‘ancestor’</td>
<td>Cham ja ‘appellation of poor people’</td>
<td>PMK 151 *ya? ‘grandmother’.</td>
</tr>
<tr>
<td><strong>kiaŋ</strong> ‘chin’</td>
<td>PC *kaŋ</td>
<td>PMK 497 *kaŋ? ‘chin, jaw’.</td>
</tr>
<tr>
<td><strong>kimuan</strong> ‘nephew, niece’</td>
<td>PC *kamuan</td>
<td>PMK 1187 *kmu ‘friend, to associate’.</td>
</tr>
<tr>
<td><strong>koh</strong> ‘cut, cut off’</td>
<td>PC *koh</td>
<td>PMK 1969 *koh ‘cut (down)’.</td>
</tr>
<tr>
<td><strong>kruat</strong> ‘citrus’</td>
<td>PC *kruac</td>
<td>PMK 846 *kruac. e.g. Old Khmer krvac</td>
</tr>
<tr>
<td><strong>kuah</strong> ‘dig, excavate’</td>
<td>PC *kuah ‘shave, scrape’</td>
<td>PMK 1881 *kus etc. ‘scratch, shave, scrape’</td>
</tr>
<tr>
<td><strong>kuat</strong> ‘gather up’</td>
<td>PC *kuac ‘gather, amass’</td>
<td>PMK 867 *kwa(a)c ‘to scratch up’.</td>
</tr>
<tr>
<td><strong>lhan</strong> ‘python’</td>
<td>PC *klaan</td>
<td>PMK 1205 *tlan.</td>
</tr>
<tr>
<td><strong>lhîoh</strong> ‘escape, loose, free’</td>
<td>PC *klaas ‘escape’</td>
<td>PMK 2064 *laas ‘to leave’.</td>
</tr>
<tr>
<td><strong>lîaŋ</strong> ‘unroll, spread out, open up’</td>
<td>PC *laŋ</td>
<td>PMK 926 *laŋ ‘to spread, to propagated’.</td>
</tr>
<tr>
<td><strong>jit</strong> ‘stupid, dull, dumb’</td>
<td>Rhade ngong ngût ‘opening mouth in exhaustion, stuttering’</td>
<td>PMK 979 *suaat etc. ‘deserted, quiet’.</td>
</tr>
<tr>
<td><strong>poh</strong> ‘to beat, hit, kill’</td>
<td>PC *poh ‘strike, pound’</td>
<td>PMK 2022 *puh.</td>
</tr>
<tr>
<td><strong>ruaŋ</strong> ‘back (of body)’</td>
<td>PC *rag</td>
<td>PMK 667 *ruaŋ[ ] etc. ‘ridge, spine, back’.</td>
</tr>
<tr>
<td><strong>sapaï</strong> ‘upper arm, sleeve’</td>
<td>PC *sapaï ‘arm’</td>
<td>Kautic *sapaï ‘shoulder’, Viet vai, Temiar pal ‘arm’: no particular source stands out.</td>
</tr>
<tr>
<td><strong>siat</strong> ‘scoop, bail out’</td>
<td>PC *sac</td>
<td>PMK 872 *sac, *saac; vowel length varies within Chamic.</td>
</tr>
<tr>
<td><strong>sinjip, siminjip</strong> ‘yawn’</td>
<td>PC *ʔaap</td>
<td>PMK 1229 *ʔaap.</td>
</tr>
<tr>
<td><strong>tuwo?</strong> ‘forget’</td>
<td>PC *war</td>
<td>PMK 1669 *war etc.</td>
</tr>
<tr>
<td><strong>wenj</strong> ‘to pedal; turn around’</td>
<td>Cham wiŋ ‘turn, whirl’</td>
<td>PMK 765 *wiŋ etc. ‘to go around, turn’.</td>
</tr>
<tr>
<td><strong>wet</strong> ‘to go around’</td>
<td>Rhade wit ‘to twist’</td>
<td>PMK 1090 *wit ‘to go round, curve, bend, turn’.</td>
</tr>
<tr>
<td><strong>wio</strong> ‘corral, stall, enclosure’</td>
<td>PC *war ‘stable, pen’</td>
<td>PMK 1669 *wiar.</td>
</tr>
</tbody>
</table>
Other Aceh-Chamic isoglosses have MK reflexes with more restricted distributions. The next group have cognates among two or more Eastern MK branches:

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>Eastern MK reflexes</th>
</tr>
</thead>
<tbody>
<tr>
<td>caga ‘Malayan bear’</td>
<td>PC *cagow</td>
<td>PMK 1817 *jkaw, Katuic, Bahnaric, Vietic.</td>
</tr>
<tr>
<td>sah ‘hollow, empty, deserted’</td>
<td>PC *sah ‘only, empty’</td>
<td>PMK 2088 *suah ‘empty’, Khmer, Bahnaric, Katuic.</td>
</tr>
<tr>
<td>toh ‘discharge, excrete from body’</td>
<td>PC *tuh ‘pour’</td>
<td>PMK 2003 *tuh ‘to pour out’, Bahnaric, Vietic.</td>
</tr>
</tbody>
</table>

Some items show particular agreement with geographically Northern MK languages, even if cognates occur elsewhere among MK:

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>Northern MK agreements</th>
</tr>
</thead>
<tbody>
<tr>
<td>kruaŋ ‘river’</td>
<td>PC *kruŋ</td>
<td>PMK 668 *ruaŋ etc. Khmu’ kruŋ agrees with Aceh-Chamic.</td>
</tr>
<tr>
<td>pa ‘to fly’</td>
<td>PC *par</td>
<td>PMK 1633 *par, Palaungic reflexes show -ar.</td>
</tr>
<tr>
<td>pet ‘reap, gather, harvest, pluck’</td>
<td>PC *pet ‘pick, pluck’</td>
<td>PMK 823 *pac ‘pick, pluck’, Palaungic reflexes show final t.</td>
</tr>
</tbody>
</table>

Strikingly, while there are Southern (Mon and/or Aslian) cognates, they are never unique to Southern MK, and show no special agreements.

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>MK agreements including southern forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>bot ‘bent backwards’</td>
<td>Cham bo? ‘to bend, curl’</td>
<td>PMK 1038 *bat; Mon, Khmer, Katuic, Bahnaric.</td>
</tr>
<tr>
<td>gulam ‘carry on shoulder’</td>
<td>PC *gulam</td>
<td>PMK 1406 *klam, *kram; Khmuic, Palaungic, Aslian.</td>
</tr>
<tr>
<td>pungo? ‘kind of owl’</td>
<td>Rhade mkao</td>
<td>PMK 1816 *ŋk[aw?], Mon, Bahnaric, Aslian.</td>
</tr>
<tr>
<td>tho ‘dry’</td>
<td>PC *thu</td>
<td>Kammu Yuan thu, Serting (Aslian) têhû.</td>
</tr>
</tbody>
</table>

The following appear to constitute loans from or into Khmer:

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>Khmer reflexes identified as likely source.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat ‘hill, rise, sloping; vertical’</td>
<td>PC *cat ‘mountain range’</td>
<td>Khmer caot ‘high, steep, sheer’, loaned into Vietic via Cham as name for Chut ethnic group.</td>
</tr>
<tr>
<td>got ‘good’</td>
<td>Cham got</td>
<td>Khmer kot got ‘just, exact’.</td>
</tr>
<tr>
<td>siaŋ ‘house’</td>
<td>PC *saaŋ</td>
<td>Khmer saaŋ ‘to build’.</td>
</tr>
<tr>
<td>som ‘wrap’</td>
<td>PC *sam</td>
<td>Old Khmer sum ‘to wind, roll, wrap up’.</td>
</tr>
</tbody>
</table>
What can the Mon-Khmer lexical borrowings in Acehnese tell us?

<table>
<thead>
<tr>
<th>Weh 'go away, leave'</th>
<th>Cham Weh 'to dodge'</th>
<th>PMK 235 *Weh 'crooked, to twist' &gt; Khmer Veh 'slip away, escape, dodge'.</th>
</tr>
</thead>
<tbody>
<tr>
<td>?Awio? 'ladle'</td>
<td>PC *?awaak</td>
<td>PMK 456 *[jwaik &gt; Khmer /vɛːk/ 'ladle, dipper' with phonology indicating other MK forms are loans from Khmer.</td>
</tr>
</tbody>
</table>

The following group are of MK or obscure origin, but there are factors to indicate (re)borrowing into Bahnaric.

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>Etyma likely (re)borrowed into Bahnaric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bagia 'shaman'</td>
<td>PC *bijow</td>
<td>PMK 1823 *[bljow]. Palaungic (e.g. Riang co?) are doubtful, thus Bahnar psjow etc. likely borrowed.</td>
</tr>
<tr>
<td>Blan 'plain'</td>
<td>PC *blan</td>
<td>Bahnar blan.</td>
</tr>
<tr>
<td>Coh 'peck, sting'</td>
<td>PC *coh</td>
<td>PMK 1987 *[ʔ]coh. Bahnaric forms do not show expected lenition *c &gt; /s/.</td>
</tr>
<tr>
<td>Hu 'ablaze'</td>
<td>Cham hu 'roast'</td>
<td>PMK 1985 *hur 'expose to heat', restricted to Bahnaric.</td>
</tr>
<tr>
<td>Linoj 'kind of eel'</td>
<td>Rhade enuj 'eel'</td>
<td>PMK 579 *[nduj] 'eel', Bahnar ranuj etc. show medial nasal.</td>
</tr>
<tr>
<td>Tiyiap 'chase, run after'</td>
<td>PC *tiyaap</td>
<td>Khmu' ngaap, Tamuon (Bahnaric) tiyaap.</td>
</tr>
<tr>
<td>Troh 'come, arrive'</td>
<td>PC *truh</td>
<td>PMK 2079 *tluuh, cf. Bahnar etc. truh.</td>
</tr>
</tbody>
</table>

Finally there is a residue of two further items.

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Takua 'neck'</td>
<td>PC *takuaj</td>
<td>No viable MK etymology is apparent, but the word form is typically MK sesquisyllabic.</td>
</tr>
<tr>
<td>Gam 'cover'</td>
<td>PC *gam</td>
<td>Khmer kaem 'cover, entrust, decorate', Proto Vietic *kæm? 'to bury' are suggestive, but do not explain initial voicing.</td>
</tr>
</tbody>
</table>

The above 50 comparisons show few indications of specific language sources, except for six isoglosses with Khmer. However, it is difficult to assess these, given the regional importance of Khmer over a millennium — they do suggest Aceh-Chamic unity until the Old Khmer period, unless independent borrowing is accepted.

### 2.2 Independent Acehnese and Chamic loans of the same or related MK roots

Listed below are 18 sets of lexical comparisons that at this stage I am interpreting as evidence of independent borrowings from MK after Aceh-Chamic separation. Some of the discrepancies between Acehnese and Chamic forms may be due to regular or other developments that I do not understand, although this may still be quite consistent with a significant period of separate development.
<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Chamic</th>
<th>MK and comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cicem</strong> ‘bird’</td>
<td>PC <em>cim</em></td>
<td>PMK 1324 *cim, <em>ciim</em> etc. MK forms suggest a long vowel, except for South Bahnaric and Aslian.</td>
</tr>
<tr>
<td><strong>dap</strong> ‘cover, put in place’</td>
<td>PC <em>dop</em> ‘to hide’</td>
<td>PMK 1261 <em>dop</em> etc. ‘to cover’. PC must have been borrowed later, after the *ə &gt; *a shift ceased.</td>
</tr>
<tr>
<td><strong>dia</strong> ‘shallow’</td>
<td>PC <em>del</em></td>
<td>PMK 1324 *kdial, <em>kfail</em>. Chamic agrees with Palaungic, e.g. Riang ‘del’ while Aceh agrees with Mon and Bahnaric.</td>
</tr>
<tr>
<td><strong>dam</strong> ‘stay overnight’</td>
<td>Rhade dam</td>
<td>PMK 1357 <em>dam</em> ‘lodge for the night, roost’. Implosive in Rhade is not expected.</td>
</tr>
<tr>
<td><strong>ginas</strong> ‘heel’</td>
<td>PC <em>kdual</em></td>
<td>PMK 1748 <em>kdul</em>. Expected reflex in Aceh would have glottal stop, forms with medial -n- are attested in Bahnaric and Aslian.</td>
</tr>
<tr>
<td><strong>glaih</strong> ‘unravel, loosen, untie’</td>
<td>PC <em>tuleh</em> ‘untie’, Cham <em>lyh</em> ‘to undo, untie, take off’</td>
<td>PMK 2067 <em>tas</em> etc. ‘unravel, unfold’.</td>
</tr>
<tr>
<td><strong>hah, haih</strong> ‘to be open’</td>
<td>PC <em>ha, ha</em> ‘open (mouth)’</td>
<td>PMK 251 *ha?, <em>hah</em> ‘to open mouth’. Vietic and Bahnaric reflexes suggest final -h.</td>
</tr>
<tr>
<td><strong>kao</strong> ‘strong, firm’</td>
<td>PC <em>khan</em> ‘hard, stiff, strong’</td>
<td>PMK 512 <em>gon</em> ‘warrior-like’. Viet <em>khang</em> &lt; Vietic/Katuic <em>kan/kaŋ</em>. Chamic might be from Khmer <em>kraŋ</em> with aspiration of -r-.</td>
</tr>
<tr>
<td><strong>kramkrom</strong> ‘sound of heavy footsteps’</td>
<td>PC <em>gram</em> ‘thunder’</td>
<td>PMK 1392 <em>gron</em> [ ] ‘thunder, rumble’. Aceh form has unexpected voiceless initial.</td>
</tr>
<tr>
<td><strong>gi</strong> ‘coals (cold), charcoal, ashes’</td>
<td>Jarai <em>anga?</em>, Roglai <em>anga</em> ‘live coal’</td>
<td>PMK 33 <em>ŋa</em>, <em>ŋa?</em> ‘live coal’.</td>
</tr>
<tr>
<td><strong>plaih</strong> ‘loosen, open up, untie, unroll, stretch oneself’</td>
<td>Rhade pleh ‘to shell corn, remove from cluster’</td>
<td>PMK 2067 <em>lo[ ]s</em>, <em>las</em> ‘unravel, unfold’.</td>
</tr>
<tr>
<td><strong>proih</strong> ‘blow, spray from mouth’</td>
<td>PC <em>prus/h</em> ‘squirt’</td>
<td>PMK 2051 *rus, <em>ruas</em> etc. Khmer <em>pruah bruas</em> ‘to squirt from mouth’, Mon <em>proh</em> ‘to squirt from mouth or trunk’.</td>
</tr>
<tr>
<td><strong>rahah</strong> ‘wash’</td>
<td>PC <em>raw</em></td>
<td>PMK 1841 <em>r[ ]aawh</em>. Vietnamese <em>rita</em> ‘to wash’. Bahnar <em>raaw</em> ‘wash rice’.</td>
</tr>
<tr>
<td><strong>wio</strong> ‘left side’, <strong>jawia</strong> ‘left-handed’</td>
<td>PC <em>ləw</em> ‘left’</td>
<td>PMK 233 *w[ ]i?. PC &lt; MK with metathesis, Aceh cf. Old Mon <em>jwiʔ</em> ‘left’, Chamic forms loaned into Bahnaric.</td>
</tr>
<tr>
<td><strong>wti</strong> ‘kind of monkey’</td>
<td>Cham <em>hawa</em> ‘macaque’</td>
<td>PMK 242 <em>swaa</em> ‘monkey’, e.g. Khmu’ <em>hwa?</em>. Riang ‘wa?’. Cham has either borrowed from MK that drops final glottal stops, or also possibly actually continues PA *[ ]uqa(Sy).</td>
</tr>
</tbody>
</table>
What can the Mon-Khmer lexical borrowings in Acehnese tell us?

2.3 Acehnese-MK isoglosses without apparent Chamic cognates

Now we come to 63 comparisons between Acehnese and MK apparently lacking Chamic cognates. I do not doubt that an exhaustive search would turn up some Chamic cognates for these, and one of the benefits of compiling a list such as this is to assist in directing such research. None the less, it is surely significant that it was not difficult to compile this substantial list of Acehnese words of MK origin that appear to lack Chamic cognates. Underlyingly it must point to a substantial period of independent development while still undergoing MK influence.

Scanning through the list below there are no stand-out dominant sources of Acehnese vocabulary. Various Mon and Khmer forms certainly appear to be sources in various cases, and strikingly there are many matches with Northern MK forms (Khasi, Palaungic, Khmuic) versus a dearth of Aslian comparisons, and the utter absence of specifically Aslian phonological or morphological features (e.g. hardening of final nasals, copyinfixation) in the Acehnese lexicon. These observations prompt a hypothesis that the historical Acehnese did not enjoy significant contact with Aslian speakers, perhaps actually remaining on the Asian mainland for a significant period of time after separating linguistically from Chamic.

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>MK comparisons and remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>biang ‘crab’</td>
<td>PMK 630 *biang ‘spider’. Widely distributed in MK, although only Khmer and Katuic show vowel agreement with Aceh.</td>
</tr>
<tr>
<td>bua ‘monkey, ape’</td>
<td>Semang Paya (Aslian) bâwe? ‘kind of monkey’.</td>
</tr>
<tr>
<td>bah ‘throw away’</td>
<td>PMK 2025 *p[oo]h ‘throw (away)’. Khmer bôh pôh ‘to throw’, Viet bô ‘to put, cast, to leave, abandon’. Khmer, Palaungic, Vietic; initial voicing suggests late borrowing from Khmer or Vietic.</td>
</tr>
<tr>
<td>cangusə? ‘frog’</td>
<td>Old Mon ngaw, Khasi jakoid. Mon most suggestive, but lack of final stop is problematic.</td>
</tr>
<tr>
<td>chuag ‘smelling of urine’</td>
<td>Khasi jîng ‘urine’. Initials disagree in voicing.</td>
</tr>
<tr>
<td>cit ‘stinging pain’</td>
<td>PMK 989 *pcat ‘bitter, astringent’. Mon phyot ‘astringent’ suggests loan form Old Mon.</td>
</tr>
<tr>
<td>Term</td>
<td>Meaning</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>crik?</td>
<td>‘tear, rip’</td>
</tr>
<tr>
<td>croh</td>
<td>‘to quench, extinguish, slake’</td>
</tr>
<tr>
<td>cumuat</td>
<td>‘term for various boils, pimples, pustules’</td>
</tr>
<tr>
<td>deh, jeh</td>
<td>‘that, there, yonder’</td>
</tr>
<tr>
<td>drop</td>
<td>‘catch, grip, trap’</td>
</tr>
<tr>
<td>galoo</td>
<td>‘hair coil, bun’</td>
</tr>
<tr>
<td>gidu?</td>
<td>‘ride, mount’</td>
</tr>
<tr>
<td>gi-</td>
<td>‘he, she’</td>
</tr>
<tr>
<td>got</td>
<td>‘to draw, tighten, pull on’</td>
</tr>
<tr>
<td>hit</td>
<td>‘slender, skinny, tight (clothes)’</td>
</tr>
<tr>
<td>jep</td>
<td>‘drink’</td>
</tr>
<tr>
<td>jo?</td>
<td>‘give’</td>
</tr>
<tr>
<td>jumah</td>
<td>‘snout, face’</td>
</tr>
<tr>
<td>kap</td>
<td>‘bite, gnaw’</td>
</tr>
<tr>
<td>keh</td>
<td>‘scrape, scratch’</td>
</tr>
<tr>
<td>khem</td>
<td>‘laugh, neigh, whinny’</td>
</tr>
<tr>
<td>kibliap</td>
<td>‘paddy, section of ricefield’</td>
</tr>
<tr>
<td>kia</td>
<td>‘in front, before, ahead, next’</td>
</tr>
<tr>
<td>kijuat</td>
<td>‘shrivelled, crumpled, shrunken’</td>
</tr>
<tr>
<td>kitap</td>
<td>‘slash, cut with whip’</td>
</tr>
<tr>
<td>kraliap</td>
<td>‘cockroach, black beetle’</td>
</tr>
<tr>
<td>krēt</td>
<td>‘gnash, gratin of teeth’</td>
</tr>
<tr>
<td>lap</td>
<td>‘to paint’</td>
</tr>
<tr>
<td>Word</td>
<td>Meaning</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>lhish</td>
<td>‘done, finished, settled’</td>
</tr>
<tr>
<td>mat</td>
<td>‘grip, grasp, control’</td>
</tr>
<tr>
<td>muan</td>
<td>‘pimple’</td>
</tr>
<tr>
<td>mup</td>
<td>‘submerged, sink’</td>
</tr>
<tr>
<td>paloh</td>
<td>‘lowlands’</td>
</tr>
<tr>
<td>prian</td>
<td>‘intestinal worm’</td>
</tr>
<tr>
<td>raq</td>
<td>‘shorter houseposts’</td>
</tr>
<tr>
<td>ridom</td>
<td>‘dark, gloomy, overcast’</td>
</tr>
<tr>
<td>ripaih</td>
<td>‘wrestle, fight, struggle’</td>
</tr>
<tr>
<td>riʔoh</td>
<td>‘sweat, perspire’</td>
</tr>
<tr>
<td>roʔo?</td>
<td>‘cut with sawing motion’</td>
</tr>
<tr>
<td>rūkoʔ</td>
<td>‘overripe (of rice so that kernels fall out when harvested)’</td>
</tr>
<tr>
<td>salop</td>
<td>‘wrapping, cover, case, wrap around’</td>
</tr>
<tr>
<td>siboo?</td>
<td>‘caress, fondle, stroke’</td>
</tr>
<tr>
<td>silop</td>
<td>‘folded/turned down, curled around’</td>
</tr>
<tr>
<td>siʔuam</td>
<td>‘warm, hot, apply heat to’</td>
</tr>
<tr>
<td>sudoʔ</td>
<td>‘take up, lift up, elevate’</td>
</tr>
<tr>
<td>tiaʔ</td>
<td>‘throw away, leave behind’</td>
</tr>
<tr>
<td>top</td>
<td>‘to stick, stab, thrust’</td>
</tr>
<tr>
<td>tuaʔ</td>
<td>‘take, accept, receive’</td>
</tr>
<tr>
<td>uaʔ</td>
<td>‘larvae, white woodworm’</td>
</tr>
</tbody>
</table>
2.4 Acehnese-Nicobar isoglosses

Thanks particularly to the efforts of Collins (1975) we can identify a modest but real body of Acehnese-Nicobarese isoglosses apparently indicative of significant linguistic contact between the respective communities. This particular result should really be the least surprising. The Nicobars are not only close to Sumatra, but are on the traditional trade routes between Sumatra, India and Burma.

Fourteen examples are listed below which compare Acehnese with Car Nicobarese. I confidently expect that we will find more if a serious effort is made to scour more diverse sources — both Acehnese and Nicobarese have under-documented dialectal diversity that deserves special attention.

<table>
<thead>
<tr>
<th>Acehnese</th>
<th>Car Nicobarese</th>
</tr>
</thead>
<tbody>
<tr>
<td>ajoih ‘burned (food), scorched’</td>
<td>ngöih ‘scald, burn, singe’</td>
</tr>
<tr>
<td>sob ‘hit, slug’</td>
<td>hōh ‘collide’</td>
</tr>
<tr>
<td>ajiam ‘toys, playthings’</td>
<td>mān-kenyum</td>
</tr>
<tr>
<td>baʔ ‘tree, branch, trunk’</td>
<td>pák ‘branch’</td>
</tr>
<tr>
<td>bisɔ ‘evil, wicked …’</td>
<td>poshi ‘to trouble, bother, disturb …’</td>
</tr>
<tr>
<td>candon ‘kind of machete’</td>
<td>ong-dông ‘cut with axe’</td>
</tr>
<tr>
<td>ciap ‘tiny (round things), pointed’</td>
<td>cǐp ‘a point’</td>
</tr>
<tr>
<td>cu ‘sea snail’</td>
<td>shōi-oi-chā ‘snail’</td>
</tr>
<tr>
<td>damḥ ‘so much, very much’</td>
<td>dammāh ‘this many, so many’</td>
</tr>
<tr>
<td>luán ‘fresh, gleaming (face, complexion)’</td>
<td>lōan ‘smooth as paper or polished surface’</td>
</tr>
<tr>
<td>mirua ‘iguana’</td>
<td>meyūha ‘snake variety’</td>
</tr>
<tr>
<td>roh ‘able, make, possible’</td>
<td>rōh ‘able, succeed, win’</td>
</tr>
<tr>
<td>wiə ‘love, affection, compassion, longing’</td>
<td>ʔuāh ‘to admire’</td>
</tr>
<tr>
<td>yua ‘order, command, tell to do’</td>
<td>yol ‘to speak, tell’</td>
</tr>
</tbody>
</table>

3 Discussion and conclusions

The proper etymological investigation of Acehnese is still in its infancy, and may well remain more or less permanently hampered by the historical complexities. However, the
What can the Mon-Khmer lexical borrowings in Acehnese tell us? 

Crudely assembled and organised data present above do provoke hints and suggestions for further research into Acehnese origins by linguistic methods. Let us take a moment to consider the following remark by Dyen (2001) in his critique of Thurgood:

If Acehnese is part of the Malayo-Chamic occupation of Sumatra, the settlements on the coast of Vietnam could have originated from there and, if they did, from northern Sumatra. Such a hypothesis would just as simply account for the “parallels” or agreements between Acehnese and Chamic. (Dyen 2001:392)

The purpose of this short paper is not to prove or disprove anybody’s particular theory, but rather to lay forth a dataset that we might objectively assess and inform future work. Yet it seems that already there is at least one striking conclusion that emerges from the comparisons listed above. The model of human migration offered by Dyen implies that the Acehnese originated as a distinct community on Sumatra, and never lived in Indo-China. Yet it is very difficult to reconcile Dyen’s model with the evidence of MK loans present here. The fact is that the Acehnese lexicon contains a significant body of loans from languages spoken on the mainland of Southeast Asia, and although there is no one standout source among mainland MK tongues, the comparisons I have compiled show a remarkable bias towards northern and eastern MK, and against Aslian. On the surface these data suggest Aceh-Chamic originating on the mainland, and even that Acehnese remained on the mainland for some time after separation before migrating to Sumatra.

On the other hand it should be pointed out that Aslian data is not particularly well represented in Shorto (2006), and this may have skewed the comparisons listed here. The Aslian languages are, after Nicobarese, the next closest MK branch geographically to Acehnese, and the hypothesis that contact with Aslian can account for characteristics of Acehnese is alive and deserving of serious investigation. The present study grew out of the availability of well organised MK lexical data provided by Shorto (2006), but that dataset is limited, and needs to be built upon extensively. In particular, the lack of a comprehensive and well organised Aslian dataset, and a Proto Aslian historical reconstruction, limits our ability to take this study further for the moment.

References


The languages of Espiritu Santo, Vanuatu

DARRELL TRYON

1 Introduction

The island of Espiritu Santo (area 4010 km$^2$), the largest island in the Republic of Vanuatu (formerly known as the New Hebrides), is located in the north of the archipelago. It has a population of 25,346 according to the most recent census (Census 1999). By now this number is estimated to have increased to approximately 30,000. The exact number of languages spoken on Santo and its offshore islands remains to be determined, ranging from 35 to 15 according to different authors.

It is generally accepted that all of the languages of Espiritu Santo are members of the Oceanic subgroup of Austronesian, and that within Oceanic they are members of the Southern Oceanic subgroup which embraces Vanuatu and New Caledonia (Lynch and Crowley 2001:20).

However, the languages of Espiritu Santo are still generally surprisingly poorly known and documented, the sole source of data for many of the languages being Tryon (1976). There is nonetheless a handful of recent grammars, including Guy (1977), Ludvigson (1989), Jauncey (1997), François (2002), Chung (2005) and Guérin (2008).

2 Previous studies

Espiritu Santo (Map 1) has a rugged mountain chain running the length of the island on the western side. To this day the definitive language map of the island remains to be drawn, not only because of the remoteness of some areas, but also because of difficulties associated with language/dialect divisions, multilingualism and the phenomenon of dialect chaining.

The mapping and classification of the languages of the island have been problematic ever since the earliest survey by Capell (1962:205), who lists only six languages, compared to the 30 speech varieties listed in the present survey.

Andrew Pawley$^1$ (1972:115) includes four languages from Espiritu Santo in his classification of the Eastern Oceanic languages. He groups them with the languages of the Banks Islands and the north-eastern New Hebrides (Vanuatu).

---

$^1$ This paper is dedicated to Andrew Pawley, whose work on the classification of the Oceanic languages has been and continues to be a source of inspiration to many.
Map 1: The languages of Espiritu Santo
Tryon (1976:80) collected 41 wordlists of 309 items from all over the island of Santo and grouped them into two sub-groups on the basis of lexical and phonological evidence. All of the languages of North and Central Vanuatu were grouped together as a subgroup of a North-Central New Hebrides group (which consisted of the languages of the Banks and Torres Islands, west and central Santo, Maewo, Ambae, north Pentecost and coastal Malakula). The languages of east Santo were grouped together as a separate subgroup.

Tryon (1976) grouped all of the languages of Vanuatu into a petal-like structure, corresponding to six equal subgroups as follows:

- North-Central New Hebrides (Vanuatu)
- East Santo
- Malekula Interior
- Erromanga
- Tanna
- Aneityum

Clark (1985) reviewed the earlier linguistic literature and concluded that Vanuatu languages fall into two major subgroups, North-Central Vanuatu and South Vanuatu (1985:219). His North-Central Vanuatu subgroup covered the same area as Tryon’s North-Central New Hebrides group, but included the East Santo and Malekula Interior subgroups, which he considered to be innovative local groups, but not separated at a higher level from their neighbours.

More recently Lynch and Crowley (2001) reassessed what is known of the languages of Vanuatu and concluded that the data available on many languages are simply inadequate to allow us to come to a clear overall picture. They feel confident, however, in confirming the existence of a North-Central Vanuatu subgroup, the Northern subgroup consisting of the languages of the Banks and Torres Islands, Espiritu Santo, Maewo, Ambae and north Pentecost, while the Central subgroup consists of the languages of the rest of Pentecost, Malakula, Ambrym, Paama, Epi, the Shepherd Islands and Efate (2001:20). No further comment on the internal structure of these subgroups was offered.

Clark (2009) revisited the languages of Santo, after short periods of fieldwork in north-west Santo (Cumberland Peninsula) between 2005 and 2009. In his exploratory paper, he says:

> I may be arguing for a ‘linkage’ connecting all the Santo languages or perhaps simply trying to elucidate a portion of the process by which the well known linguistic diversity of Vanuatu has emerged from some state of relative initial unity — a process which, it seems more and more clear, has taken place largely in local terms and on local ground. (Clark 2009:1)

In other words he sees Espiritu Santo as an ‘arena’ in which various innovations arise, extend themselves and interact within a linguistic complex, more or less self-contained, and with only occasional interaction with other places. Clark examines some regular consonant changes on the basis of his reconstruction of Proto North Vanuatu, many of which cut across lexically based groups. He then discusses a number of lexical and two pronominal innovations which appear in nearly all Santo languages, and only occasionally in small areas of neighbouring islands. Clark does not seek to draw any firm conclusions, especially as so little grammatical data is available for the languages of Santo.
3 The current language situation on Santo

The principal aim of this paper is to attempt to plot the numerous Santo speech varieties as recognised by the people of Santo themselves (see Map 1).

It has only been with the considerable assistance of the ‘fieldworkers’, the ni-Vanuatu researchers attached to the Vanuatu Cultural Centre, that the present map has been compiled over the past three years. However, it should still be regarded as provisional, although hopefully a significant improvement on previous attempts at mapping the languages of this island.

Perhaps the most frustrating difficulty in the course of this research has been the variation in language and dialect names used and recorded in the literature on Vanuatu languages (see Tryon 1972, 1976, 1979). In this paper, currently accepted language/dialect names are used, variant names being indicated in footnotes where appropriate.

There are 33 extant languages, and a multiplicity of dialects spoken on the island of Espiritu Santo. Two languages, Aore and Nethalp, have become extinct in recent decades, while another ten languages/dialects are remembered as having been spoken on Santo within living memory.

Of course a major problem is to distinguish between language and dialect, defined as lanwis and smol lanwis respectively in Bislama, the English-based pidgin of Vanuatu. Ni-Vanuatu, as Vanuatu nationals are called, tend to over-differentiate, each dialect bearing a separate name. For purposes of mapping the languages of Santo, the criterion of mutual intelligibility as expressed by the local fieldworkers has been invoked as far as possible, although it is likely that the number of distinct languages spoken on the island will be reduced as research proceeds. There is also the phenomenon of dialect chaining, extending from the south-west/central area of Espiritu Santo up to Big Bay (Bay of St Philip and St James) and on up to Cape Cumberland. The languages of this part of Espiritu Santo are very closely related in any event, often leaving the demarcation between language and dialect uncertain. Added to this is a strong tradition of multilingualism, fostered by exogamous marriage.

More significant in the case of Santo perhaps are the extensive population movements which have taken place there over the past century, even more marked in recent times. The major movement has been from the interior of the island to the coasts, north, south, west and east, both in order to access European trade goods, and because of the implantation of Christian missions. What is characteristic of Santo populations is that many of them have moved in more than a single direction. Indeed many of the language groups of central Santo have moved in several directions at once, a veritable explosion, always towards the coastal areas some distance from their home base, often following the course of some of the major rivers. Added to this is the unusual tradition in much of Santo of relocating villages following the death of a village resident, often involving a new village name (Guiart 1958).

A second complicating factor springs directly from this, namely the difficulty in locating the same place names, especially village names, on official maps, from decade to decade, to the point that many of the village names which appear on the IGN maps of Espiritu Santo published in the 1960s and 1970s are unrecognisable today, especially for villages located in inland areas.
Since Vanuatu gained independence in 1980, there has been a major change too in terms of the urban and peri-urban areas around Luganville, the major town of Espiritu Santo, as large numbers of ni-Vanuatu migrants from other islands have moved to Santo. Indeed the whole of the south-east of the island, from Shark Bay right around to Tangoa in south Santo, formerly lightly populated, has become a veritable melting pot, with migrants from as far south as Tanna, affected by population pressures on their home islands, and seeking economic advancement.

The map of the languages of Santo presented here largely represents traditional home areas, including the urban and peri-urban areas around Luganville, where the original populations were forced to retreat towards the interior of the island as colonial and urban pressures built during the period of the Condominium (1906–80) and particularly during the major upheaval caused by the huge American presence during World War II, as Santo became a forward base during that conflict in the Pacific.

The language boundaries marked in Map 1 are approximate only and have no value in terms of traditional land ownership. They simply encircle villages in which a particular language is known to be spoken today, or was formerly spoken. Uninhabited areas or areas not recognised as belonging to any particular language group remain unmarked. Of course this situation will inevitably be subject to modification as research continues.

While the map of the languages of Santo presented here is intended in the main to represent traditional home areas for the different languages, it must be remembered that in many villages around the island today, and indeed in a large number of Vanuatu villages, a multiplicity of languages may be spoken, both as a consequence of population movement and displacement and exogamous marriages, as well as migration to Espiritu Santo from many other islands of the Vanuatu archipelago.

The languages of Espiritu Santo currently recognised by the Santo-based fieldworkers of the Vanuatu Cultural Centre are as follows:

<table>
<thead>
<tr>
<th>Language</th>
<th>Villages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>North-West Santo area</strong></td>
<td></td>
</tr>
<tr>
<td>1. Valpei (Tavanlav)</td>
<td>Wunpuko, Valpei, Petani, Matalip, Pwar, Molpoe, Hakua, Maroa, Pwat, Pwatmwel (Onmwertev), Mwalovuko, Wunapak [300]</td>
</tr>
<tr>
<td>2. Nokuku (Vevatot)</td>
<td>Olpoe, Nokuku, Lajmoli, Vunon, Penaoru, Petawota [250]</td>
</tr>
<tr>
<td>3. ‘Oa</td>
<td>Tasmate, Sulesal, Vasalea [300]</td>
</tr>
<tr>
<td>4. Vunapu</td>
<td>Vunapu, Pesena [250]</td>
</tr>
<tr>
<td>5. Piamatsina</td>
<td>Piamatsina, Piamota, Peavot [250]</td>
</tr>
<tr>
<td>6. Tolomako (Big Bay)</td>
<td>Jureviu, Tuturu, Tavunamalo, Peavot, Vasi, Pialulup, Matantas [900]</td>
</tr>
</tbody>
</table>

2 Approximate numbers of speakers indicated in square brackets.
3 Vunapu and Pesena are part of the same community (Clark 2006), Vunapu being Seventh-Day Adventist and Pesena Catholic.
4 Piamatsina speakers are also reported at Penaoru and Petawota on the west coast.
5 Vasi and Pialulup speak a separate dialect of Tolomako (Clark 2006), with a /k/>/l/ shift.
<table>
<thead>
<tr>
<th>Language</th>
<th>Villages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>West Santo area</strong></td>
<td></td>
</tr>
<tr>
<td>7. Kula$^6$</td>
<td>Wusi, Kerepua, Elia 1 [350]</td>
</tr>
<tr>
<td>8. Bura$^7$</td>
<td>Linduri, Putonro, Saktu, Maram, Patiare, Mapten, Nukupospos [300]</td>
</tr>
<tr>
<td>10. Akei$^9$</td>
<td>Toramaori, Lalaolo, Tovotovo, Kerevinumbu, Tasiriki, Ukoro, Malovira, Tasmalum [4000]</td>
</tr>
<tr>
<td>11. Daruru</td>
<td>Pelmol [100?]</td>
</tr>
<tr>
<td><strong>South Santo area</strong></td>
<td></td>
</tr>
<tr>
<td>12. Retatur</td>
<td>Tanovusivusi [100]</td>
</tr>
<tr>
<td>13. Ale</td>
<td>Fimele, Wailapa [500]</td>
</tr>
<tr>
<td>15. Ande$^{10}$</td>
<td>Tanmet, Lotunai, Ipayato, Tasmalum, Morouas [500]</td>
</tr>
<tr>
<td>16. Araki</td>
<td>Araki Island [10]</td>
</tr>
<tr>
<td>17. Movono</td>
<td>Tangoa Island [370]</td>
</tr>
<tr>
<td>18. Farsav$^{11}$</td>
<td>Nambel, Tanovoli, Narango, Funafosi, Nambauk [400]</td>
</tr>
<tr>
<td><strong>Central Santo area</strong></td>
<td></td>
</tr>
<tr>
<td>19. Tiale/Merei</td>
<td>Tavuimoli, Nazaraka, Mataipevu, Patuitano, Morokari, Angoru, Tombet, Navele, Vusvogo [400]</td>
</tr>
<tr>
<td>20. Kiai</td>
<td>Wailapa, Fortsenale, Namoru [450]</td>
</tr>
<tr>
<td>21. Ko/Mores (Farmores)</td>
<td>Tanmet, Lemben, Tsarailan, Namafun, Patmarifu, Sarete, Mavunlif [200?]</td>
</tr>
<tr>
<td>22. Moiso</td>
<td>Moriuli$^{12}$ [100]</td>
</tr>
<tr>
<td>23. Toksiki/Soisoru</td>
<td>Morkriv, Pilnuri, Bengie [200]</td>
</tr>
<tr>
<td><strong>East Santo area</strong></td>
<td></td>
</tr>
<tr>
<td>24. Tamambo$^{13}$</td>
<td>Malo [4000]</td>
</tr>
<tr>
<td>25. Aore$^{14}$</td>
<td>Aore [EXTINCT]</td>
</tr>
</tbody>
</table>

---

6  Fieldworker Frank Lengki reports that there are four extinct languages in the Wusi area, as follows: Latu, Mwe’ea, Sinie and Menie. Nothing is known of these languages.

7  Bura is reported to be intelligible as far north-east as the Big Bay bush area.

8  Spoken south of the Pareo River.

9  Akei is perhaps best thought of as a dialect chain.

10 Spoken between the Buvo and the Navaka Rivers.

11 Recorded as in Tryon (1976) as Nambel.

12 Nokanoka, today extinct, is reported to have been formerly spoken at Moriuli.

13 Fieldworkers Vira Joseph and Tom Rasu report that Reti Maranjino, formerly spoken in north-east Malo, became extinct in the 1950s.

14 The last Aore speaker died in the 1970s. See Tryon (1976) for lexical information.
26. Biliru\textsuperscript{15} Tambotalo, Belnatsa, Beleru, Belembut, Lambue [3]
27. Farafi\textsuperscript{16} Butmas, Tur, Maniok, Shark Bay, Naturuk [300?]
28. Ngen Vanafo (Tanafo),\textsuperscript{17} Mon Exil, Palon, Shark Bay\textsuperscript{18} [250]
29. Ati (Meris/Miris) Fumbak, Naturuk, Nambauk [85]
30. Se Vanafo, Butmas [20]
31. Atin (Farnanatin) Nambauk (Patunfarambu),\textsuperscript{19} Fumatal [120]
32. Farnanto Nambauk, Tanmet, Tafua [100]
33. Mavea,\textsuperscript{20} (Lonavu) Mavea,\textsuperscript{21} Aissi, Tutuba [500]
34. Nethalp\textsuperscript{22} Lorediakarkar [EXTINCT]
35. Nekep\textsuperscript{23} Hog Harbour,\textsuperscript{24} Port Olry, Kole 1 [4000]

4 Conclusion

Currently there are 33 languages/speech varieties recognised by the people of Espiritu Santo. Data available for all but a handful are insufficient to allow any definitive statement about their classification and subgrouping apart from the consensus that they are all members of the Northern Vanuatu subgroup. Based on tentative shared phonological and lexical innovation, they all appear more closely related to each other than to languages from any island outside Santo and so to constitute a Santo subgroup within Northern Vanuatu.

\textsuperscript{15} Recorded in Tryon (1976) as Tambotalo, the original language of Lugarville.
\textsuperscript{16} Recorded in Tryon (1976) as Butmas.
\textsuperscript{17} Ngen speakers originated at Tanafo and Palon, moved to Lataro Island (Pilot Island) and later resettled at Natawa, Shark Bay and Maniok. Ngen was recorded in Tryon (1976) as Shark Bay 1. The original language of Pilot Island, Nitro, is extinct.
\textsuperscript{18} Lataroa Island is uninhabited.
\textsuperscript{19} Patunfarambu is a Mission centre, hosting a wide variety of languages from surrounding areas.
\textsuperscript{20} The late Kuvu Keven reports that people from Mavea moved to Tutuba in 1946, after World War II.
\textsuperscript{21} Guérin (2008) reports 210 people on Mavea, of whom only 34 are fluent speakers; also 100 at Deproma on Santo mainland, of whom 10 are fluent speakers. Aissi is uninhabited today.
\textsuperscript{22} Recorded in Tryon (1976) as Lorediakarkar.
\textsuperscript{23} Also known as N’kep and Sakao (Guy 1977). The late fieldworker Kuvu Keven reports that there were three dialects of Nekep spoken between Cape Quieros and Shark Bay, as follows: N’hemel (northern dialect), N’kep (southern dialect), Nethalp (extinct, but formerly spoken at Lorediakarkar, Shark Bay area), today inhabited by migrants from other islands. At Port Olry (Port Olry 1), the inhabitants are Tolomako speakers from Big Bay, relocated to Port Olry by Catholic missionaries during the 1950s. Lelek and Loren are inhabited by people from Pentecost, Lorediakarkar by people from Paama and Ambrym, while in villages such as Nomanio, Pelron, Lathhi and Tontas Bislama is the main lingua franca. Pilot Island (Lataro) is inhabited by people from Vanafo, settled there by missionaries, as is Natawa village.
\textsuperscript{24} Also Sarah 1, Sarah 2, Sarah 3, Kole 1, Kole 2, Levan, Lovalikar, Luarev, Bne, Lorevalk, Luat, Lovocar, Lorim, Hala and Rotal. Today Kole 2 is inhabited by people from Malakula and Ambae, making Bislama the main intergroup language.
References


Ludvigson, T., 1989, A Kiai wordlist. Working papers in Anthropology, Archaeology, Linguistics, Maori Studies, No.81, Department of Anthropology, University of Auckland.


1 Introduction

In his sweeping 1981 article on the pattern of diversification and subgrouping in Oceanic, Andrew Pawley wrote, ‘It is not the case that there is one law for Polynesians and another for Melanesians’. 1 In that article, Pawley proposed that the development of the different ‘branches’ of Oceanic in north-western Melanesia is ‘best explained by assuming a fairly swift dispersal of Proto Oceanic speakers over most of this region, followed by the gradual breakup of the dispersed language community into smaller units corresponding generally to natural geographic zones within which regular communication could be easily maintained.’ This paper is an investigation into Pawley’s claim relative to Hawaiian and northern East Polynesia. 2

Specifically, I will investigate lexicon uniquely shared by Hawaiian and Tuamotuan, a frequently overlooked language of northern East Polynesia. A set of terms uniquely shared between Hawaiian and Tuamotuan gives the appearance of reflecting innovations which took place in a unique common lower-order protolanguage. The existence of this set of 45 ‘Tuamotuic’ terms competes with 31 ‘Marquesic’ terms included in Biggs and Clark (1992, henceforth POLLEX), and listed by Marck (1996) as distinctive innovations of a Marquesic subgroup which includes Hawaiian. An investigation of Marck’s 31 terms reported here leaves only seventeen uncontested. Thus under a strict subgrouping scheme, Hawaiian could as easily be ‘Tuamotuic’ with Marquesic borrowings as, according to the widely accepted proposal by Green (1966), Marquesic with Tahitic borrowings.

In addition to terms shared exclusively by Hawaiian and Tuamotuan, I provide a similar list of terms shared exclusively by Hawaiian, Tuamotuan and Tahitian, another list of terms uniquely shared by Hawaiian, Tuamotuan, and Marck’s (1996) Proto Tahitic, and finally a small list of terms shared exclusively by Hawaiian, Tuamotuan and Marquesan. All four lists provided here, plus the adjusted set of innovations exclusively shared by Hawaiian and Marquesan, display regular consonant correspondences, rather than irregular

---

1 It is a great pleasure to be included in this volume honouring Andy Pawley, a teacher whose standards in research, teaching, and aloha have had a profound effect on my work with the Hawaiian language and its revitalisation.

2 Appendix 1 lists the languages cited and abbreviations used for them, and places them in large subgroups. Appendix 2 gives regular consonantal reflexes of Proto Polynesian in those languages.
correspondences which would indicate late borrowing. Data from Mangarevan, a Marquesic language, are also included, but with caution, due to the high likelihood of borrowing between Mangarevan and Tuamotuan, given the location of Mangareva at the end of the Tuamotu archipelago.

While uniquely shared vocabulary exhibiting regular phonological correspondences can potentially be explained as either shared retentions or parallel innovations, the likelihood of such explanations being correct diminishes as the number of uniquely shared vocabulary items increases. A more difficult alternative explanation to eliminate is extensive borrowing. If the languages involved in the borrowing are closely related and have the same, or highly similar, phonologies and grammars, detecting borrowing is extremely difficult. The more closely related the borrowing languages, and the more extensive the borrowing, the more difficult it is to characterise the two speech communities as different languages rather than different dialects. Indeed, the proposal here is that Hawaiian developed as part of an early East Polynesian dialect chain. This scenario is exactly what Pawley (1981) described as characteristic of north-western Melanesia and what he (1996) later described in more detail for Polynesia, especially western Triangle Polynesia.

The implication of the data presented later below is that Hawai`i was settled earlier than generally accepted, that is, Hawai`i was settled before distinctive Proto Marquesic and Proto Tahitic languages developed from Proto Central East Polynesian. The indications of a specific relationship of Hawaiian to languages characterised as Tahitic or Marquesic can best be described as due to diffusion resulting from a long period of contact between Hawaiian and other languages of northern East Polynesia.

2  East Polynesian subgrouping

The standard subgrouping of East Polynesian has it as a well defined subgroup of Nuclear Polynesian descended from a distinct Proto East Polynesian interstage language (Elbert 1953; Pawley 1966). Proto East Polynesian then divided into Rapanui and Central East Polynesian (Green 1966:17–18) followed by a subsequent split of Proto Central East Polynesian into Tahitic (Green 1966:26–28) and Marquesic subgroups (Green 1966:18–26).

Green’s Marquesic includes Mangarevan, Hawaiian and what Green sees as two Marquesan languages: Northern and Southern Marquesan. Green’s Proto Marquesic splits initially into Mangarevan and the ancestor of all other Marquesic languages. This is followed by a split into Northern Marquesan and Proto Southern Marquesan-Hawaiian, with a final split of Hawaiian from Southern Marquesan.

Marck (1996) provided support for Hawaiian as Marquesic with additional data from POLLEX. Marck differed from Green, however, in proposing that Mangarevan diverged from Marquesic after, rather than before, Hawaiian. Marck’s (1996:503–504) proposal is that Proto Marquesic split first into Hawaiian and Proto Nuclear Marquesic, the latter being the source of the two Marquesan dialects and Mangarevan.

There has been some reinvestigation of the position of Rapanui and Mangarevan within East Polynesian. In responding to a proposal by Langdon and Tryon (1983) that Rapanui included both ‘Futunic’ and East Polynesian strata, Clark (1983) showed that ‘Futunic’ elements could be ascribed to retentions from Proto East Polynesian, but that other evidence indicated a late genetic relationship of Rapanui with Marquesan and Mangarevan. In response to a proposal by Fischer (2001:119) that Mangarevan subgrouped with Rapanui, not Marquesan, but had a Marquesic superstratum, Marck (2002:229) showed
that data provided by Fischer actually supported subgrouping of Mangarevan within Marck’s ‘Nuclear Marquesic’.

3 Hawaiian and Marquesic

3.1 Marquesic lexical innovations in Hawaiian

Our focus here is on terms in Hawaiian that could have been directly inherited from Proto Marquesic. Marck’s (1996) proposal that the initial split in Marquesic was between Hawaiian and all other Marquesic languages is followed here. In listing 31 Proto Marquesic lexical innovations from POLLEX, Marck (1996) noted that Biggs dismissed four of the six initial Proto Marquesic innovations proposed by Green (1966). The rejected terms were *(nga)ngahu ‘bite’, which is also reflected in non-Marquesic languages, *pi ‘full’ which lacks a likely Hawaiian reflex, *rongofulu ‘ten’ which also lacks a Hawaiian reflex, and *wai-mata ‘tears’ for which Tahitic reflexes also exist. Below I discuss fourteen of the remaining 29 Proto Marquesic terms in POLLEX which I reject as I argue that they do not have a directly inherited Hawaiian reflex. Thus only seventeen of Marck’s 1996 innovations shared by Hawaiian and Marquesan remain uncontested as potential directly inherited Hawaiian reflexes of Proto Marquesic innovations.

1. *koo-mata, listed by Marck without a meaning and without a Hawaiian cognate, must be eliminated because of Tahitic cognates.
   Mqa komata ‘part of fruit that holds the stem’, Mva komata ‘indentation in fruit that holds stem’ has a Tahitic cognate in Tua kōmata ‘the cap or base which holds a tiny coconut (pūriri) to the stem’ (Anaa, Hao, Reao). A possible Hawaiian cognate is found in ‘ōmaka waiū ‘nipple’, but cognates with this meaning also occur in Tahitic languages (Mao, Rar kōmata).

2. *koo-piko ‘youth, juvenile person’ is reconstructed with penultimate *k, on the basis of Mqa kopi‘o ‘not yet fully mature’. The -k- is reflected neither in Haw ʻōpio ‘a youth’, ʻōpiopio ‘young of any living thing including immature fruit’, nor in Mqa kopio, ʻopio ‘not mature of fruit’, the entry in Le Cléac’h’s 1997 dictionary. The conflicting forms in POLLEX and Le Cléac’h can be explained by the sporadic epenthesis of the reflex of PEPn *k in penultimate position in Central East Polynesian languages, also found in Mao maoa, maoka, Haw mo’a from PEPn *maoa ‘cooked’ and in Tua tara mea, tara meka ‘various red spiny sea creatures’. The reconstruction is thus provisionally modified to *koo-pio.
   Tua topio (possibly tōpio) ‘immature fruit of the pandanus’ (Takume, Raroia) appears cognate. The irregular initial consonant correspondence between Haw ʻōpio and Tua topio can be explained by a sporadic conditioned change of *t to *k

---

3 The introduction to POLLEX 1995 states that it contains 38 Proto Marquesic reconstructions, an increase of seven over POLLEX 1992. For this article, no effort was made to look through POLLEX 1995 for those seven.
4 Geraghty (pers. comm.) notes that Green’s *pi has cognates in Tūvalu and Taku pī ‘container (full)’ and possibly also in Wayan Fijian bī ‘many’ and Standard Fijian bī ‘heavy’.
5 Stimson (1964) lists island names for dialects of Tuamotuan, a practice followed here.
6 Marck (2000:70) noted this phenomenon as epenthesis of a glottal stop in Marquesan, but did not correct the POLLEX reconstruction.
in the environment *tV(C)V(C)VCV. This change is most clearly evident in Hawaiian in the following doublets: Haw kā'ili, ā'ili ‘to gasp, labour for breath’; Haw kānana, ānana ‘to strain’; Haw kūlepe, ālepe ‘harelip’. The rule also applies sometimes in Marquesan and Tuamotuan: Mqa ‘enata, ēnana, kenana from PPN *taangata ‘human beings’; Tua kainoka (Vahitu, Anaa) but Tah rai no’a ‘a parasitic vine’; Tua kakariuri (Vahitu) from PPN *talitali-quli ‘remora fish’. In addition the following Hawaiian–Tuamotuan pairs are affected by this rule: Tua tarakihi ‘fish with very sharp dangerous spines’ (Anaa), Haw ‘ala’ihi ‘squirrel fish, which is noted for its spines’; Tua tūruta ‘extensive assemblage, collection, great heap, crowd’ (Anaa), Haw ‘āluka ‘large group, herd as of goats’; Tua tūtururū ‘to ruffle up, spread out the feathers’ (Vahitu), Haw kū’ululū ‘shivering with cold as a bird in the rain’.

This leads us to further modify the reconstruction to *too-pio, but the fact that it is reflected in Tuamotuan means that it must be rejected as a Proto Marquesic innovation.

3 *mio ‘extinguished’ is reflected by Mqa mi-mio ‘fade’, Mva mio ‘extinguished’ and allegedly Haw mio ‘disappear quickly, move swiftly as a stream of water, make off quickly, steal, wilt, depart quickly, current’. The Marquesan, Mangarevan, and Hawaiian terms are not identical in meaning. Furthermore, they can be individually derived both semantically and phonologically from either of two forms PNPN *miomio ‘whirlpool’ and PPN *qamio ‘variable (of direction of motion)’ that existed before Proto Marquesic and which have reflexes in other East Polynesian languages.

4. *paka-tai ‘salt’ (Haw pa’akai) is reflected among Tahitic languages in Tah pa’a tai ‘salt’, a rare alternative to miti pōpa’a ‘salt’ (lit. ‘Caucasian sea water’) and is thus not a Marquesic innovation. In any case, the term Mqa, Tah pa’a tai is also suspect as a post contact borrowing from Hawaiian. Ancient Hawaiian food was distinctive in its heavy use of salt that was produced in salt pans. During the early contact period Hawai‘i was an important source of salt for sailing ships, on which Hawaiians and other Polynesians often served as crew. A number of post European-contact loans from Hawaiian occur in French Polynesia. An especially obvious one is Tah, Mqa, Tua (no dialect information) tūtaepere ‘sulphur’ cf. Haw kākaepele ‘sulphur’ (lit. ‘magma excrement’).

5. *maatuke ‘dark brown, dark in colour’ is reconstructed based on Haw māku’e ‘brown of skin’ and Mqa matu’e, matuke ‘brown of skin’. Mqa matu’e also means ‘sea urchin’, alternating with fetu’e and hatuke. PPN *watuke (> PEPN *fatuke > Haw hāku’eku’e, hā’uke’uke, hā’ue’ue) refers to the slate pencil urchin, which in Hawaiian waters has a distinctive red to reddish brown colour. Brown, especially as a skin colour, is seen in Hawaiian as a form of red.

The external Kapingamarangi maduge ‘slate pencil urchin’ suggests borrowing in the history of the innovation *maatuke. The borrowing may have been from Kapingamarangi into either Hawaiian or Marquesan or both, and thus *maatuke must be eliminated as a uniquely shared innovation. A genetic connection between the Northern Outliers and East Polynesian (Wilson 1985, 2008) might also have been followed by continued sporadic contact between the two areas.
6. *tokorau ‘north, northerly wind’, for expected *tokerau from PCEPn *tokelau, reflected in Hawaiian, Marquesan and Mangarevan, is also found in the Northern Outliers: Takuu *tokorau ‘north, northerly wind’, Luangiua *kolu ‘north’, suggesting participation in the borrowing mentioned in 5.

7. *ra(a)ngahu ‘charcoal’ is suspicious as the primary distinguishing feature shared by Marquesan, Hawaiian and Mangarevan is an apparent metathesis from an earlier PNPN *ngalafu. Interaction between n and l, including metathesis, is extremely common in Hawaiian and thus Haw lānahu (also nānahu) ‘charcoal’ may be an independent parallel innovation.

Other examples of such interchange of Haw l and n are Haw nalo < PPN *lango ‘fly’ (cf. also Mao ngaro, rango ‘fly’), Haw wānana, wālana ‘prophecy’ < PEPn *waananga ‘special pronouncement’ (my reconstruction), Haw nalu ‘amniotic fluid’ < PPN *lanu ‘bathe or wash in fresh water’.

Casting further doubt on this as a shared innovation is the frequency of interaction between *n and *r (or in the case of Marquesan between n and ‘ < PEPn *r) in northern East Polynesian, e.g. PEPn *manufiri ‘visitor’ > Tah manufiri, manuhiri, manuhini, manihiri, manihini, Haw malihini, Tua marihini (Vahitu, Anaa), manihini (Fangatau), marihinui (Napuka), Mqa manihi’i, ma’ihi’i.

8. *kena ‘glowing, red hot’: Haw ‘ena ‘red hot’, ‘anger’, Mqa kena ‘red hot’ have a Tuamotuan cognate in Tua kega ‘red, red hot, flushed red’ (no dialect information), thus supporting an earlier *kena and excluding the possibility that this is a Marquesic innovation.

Both the Tuamotuan and Hawaiian terms are associated with an earlier PEPE *r (Tua reka ‘firebrand’ (Tatakoto), Haw iō-‘ena, iō-len ‘wild, savage’) suggesting a late borrowing from Marquesan into Hawaiian and Tuamotuan after PEPE *r and *k merged as glottal stop in Marquesan, the glottal stop alternating with Mqa k through hypercorrection.

9. *naakuu ‘kind of plant’: Hawaiian reflex nāū forms a doublet with nānū ‘native gardenia’. The latter supports deriving the glottal stop of Mqa na’u from an earlier PEPE *r that then alternated with n as described under 7. Terms where PEPE *r is reflected as Tua k and Hawaiian glottal stop are strongly suspected as late borrowings from Marquesan as discussed under 8.

10. *ketaha ‘fern species’ with the doublet Haw ‘ēkaha, ‘ākaha is derived in POLLEX from PPN *katafa ‘bird’s nest fern’. Marck (2000:83–86) discussed low vowel raising in antepenultimate syllables as something that ‘has to a large extent operated independently in each language in which it has occurred.’ This observation greatly lowers the value of this correspondence between Hawaiian and Marquesan as a possible shared innovation. The existence of a doublet in Hawaiian is also a problem for a proposal that the terms are inherited from Proto Marquesic.

11. *kaihue ‘steal’ has a likely cognate in Tua kaihūa ‘to eat greedily, voraciously, glutonously’ (no dialect information), excluding the possibility that this is a Marquesian innovation.

Stealing is also related to eating in PPN *kai-haqa ‘steal’ (lit. ‘forbidden eating’). The likely derivation of Tua kaihūa from PPN *kai ‘eat’ + *fua ‘only, merely’ is
semantically consistent with the Proto Polynesian term in that both mean eating without permission. Haw ‘aihue ‘steal’ thus appears to derive from an innovation shared by Tuamotuan, Hawaiian and Marquesan.

The raising of the final vowel of the sequence uc(C)a is found in other Hawaiian terms, e.g. Haw huehue < PPn *fuafua ‘pimple’ and in the doublet Haw huna, hune ‘fragment’. The change ua > ue is characteristic of other terms shared among Tuamotuan, Hawaiian and Tahitian discussed below in §4 under 1 and 11, and in §5.2 under 1.

12. *tokete ‘brother-in-law’, reflecting PEPn *taqokete, is another term where the shared innovation involves a common sporadic phonological change in Hawaiian, Marquesan and Mangarevan, also reflected in non-Marquesic Penrhyn tōkete. The phonological change also affects other East Polynesian languages, e.g. PNPn *fungaona ‘child-in-law’ > Mao hungaona, hungōna, Tah hungō’a. The change has variable application in Hawaiian and is likely to have spread to this term through internal lexical diffusion in each of the languages in which it occurs.

13. *pana ‘bow’ was noted by Marck (1996:502) as occurring in various East Polynesian languages with ‘various meanings assigned, usually having to do with the releasing action associated with bows and sling shots, recoiling, bouncing and so forth’. He also noted that the use as a name for the bow in Hawaiian and Marquesan is ‘a fairly natural semantic shift’. Its occurrence as a nominal in other East Polynesian languages (an in Tūvalu pana ‘slingshot’) weakens its value as an exclusively shared innovation of Hawaiian and Marquesan.

Furthermore, pana is apparently an irregular development of PPn *fana ‘bow’, reminiscent of other irregular developments of PPn *f to p and w in East Polynesian. Green (1966) notes that Mva na’a ‘bow’ retains the older form PPn *fana with metathesis. Accepting Marck’s (1996) subgrouping of Mangarevan with Marquesan, Marquesan pana is either a late parallel innovation with Hawaiian or a borrowing between Marquesan and Hawaiian.

If the distribution of pana ‘bow’ is due to borrowing, Hawaiian is the more likely source than Marquesan. Hawaiian shows a number of unique irregular reflexes of PEPn *f as p, which are usually not shared by other East Polynesian languages, e.g. Haw peopeo ‘round, globular’, pō-heoheo ‘knob’ < PPn *feo ‘coral’ (cf. Tah feo ‘a kind of coral that irritates the skin’); Haw penopeno ‘dirty, smutty’ < PCEPn *fengo ‘bad smell’ (cf. Mao whengo ‘fart’); Haw pio ‘whistle, typically with fingers inserted in mouth’ < PNPn *fio ‘whistle’ (cf. Mao whio ‘whistle’); Haw pā ‘a bound clump as of grass’, Mao huka-huka ‘lock of hair’; Haw lohelohoe, ʻō-lopelope ‘dragonfly larvae’; Haw ʻūpā ‘open and shut, tools that open and shut, e.g. scissors, tongs, bellows’, ʻūhā ‘thigh’ < PCP *kuufaa ‘thigh’ (cf. Mao kūwhaa ‘thigh’); Haw pai-o ‘combat’ < PPn *fai ‘fight’ (cf. Mao whawhai ‘fight’).

A related phenomenon bearing on the position of Hawaiian is that Hawaiian p sometimes reflects PEPn *w, or perhaps *f derived from PEPn *w, as in Haw lepa ‘flag’ < PEPn *rewa ‘flag’ < PNPn *rewa ‘free float in air or on water, be suspended’; Haw lepelepe ‘wattles’ < PPn *rewerewe ‘that which is hanging down and ragged’. Sometimes the reverse change is found: Haw walu, wolū ‘oilfish’ < PNPn *palu.
Marck (1996:498) notes that Proto East Polynesian has been reconstructed as having replaced PPN *f with PNPN *h before *u and *o, yet unique Hawaiian terms with *pu for PNPN *fu suggest that early Hawaiian had the sequence fu-, later replacing it with either hu- or pu-. Note for example: Haw punapuna ‘mealy, as taro or sweet potato, scattered, pulverised’, hunahuna ‘crumbs’ < PNPN *funga ‘crumb, grain, or scrap from crumbling or scraping’ (cf. Tah hu’a ‘particle’, Mqa huka ‘piece, bit’); Haw pua-na ‘opening or main line of a chant or song’, ‘pronounce’ < PPN *fua ‘commencement of a chant or song’ (cf. Tua hūa ‘the principle portion or body verse of the native chants, to commence to chant the hūa’ [Amanu, Anaa], Mqa hua ‘refrain of a song’); Haw pue ‘feathers just above the tail of a bird’, puapua ‘tail feathers of a bird’, Tongan fuefue ‘tail and rump of a roast pig’, Mao hua ‘buttocks’; Haw pua ‘baby mullet’ < PPN *fau(fua) ‘fish species, mullet (-like)’.

In contrast to the PNPN *fu sequence, Hawaiian reflects the PNPN *su sequence as hu e.g. PNPn *suu ‘fart’ > Haw hū ‘roar of wind’, PPN *sua ‘turn over, lever up as soil with a stick when weeding’ > Haw hua-i ‘uncover earth oven’, PNPN *sue ‘puffer fish’ > Haw ‘o’opu-hue.

Hawaiian also commonly reflects PNPN *f as Haw w, not only initially before another PNPN *f as is the case in Tahitic languages, e.g. PNPN *fafa ‘mouth’ > Haw wahā; PNPN *fafine ‘woman’ > Haw wahine; PNPN *fāfo ‘outside’ > Haw waho, but also in words where Tahitic languages retain PNPN *f as *f. Examples include: Haw wahī ‘to wrap’ < PEPN *fa-fii ‘wrap’ (reconstruction mine), perhaps a conflation of PPN *fiifii ‘package of fish’ + *qafi-hi ‘wrap up a parcel, a parcel of food’ (cf. Tongan fīfī ‘enclose fish in a plaited coconut leaf’, Kapingamarangi hīhī ‘package, wrap’, Rap hahi ‘to wrap’, Mqa fafī, hahi ‘to wrap’); Haw ā-wihi maka ‘wink’ < PPN *fīfī mata ‘eyelashes’ (cf. Tūvalu fīfī-mata ‘eyelashes’, Nukuoro hīhī mada ‘eyelashes’, Rap hihi ‘eyebrow’, Tah hīhīmata ‘eyelashes’); Haw ‘awa’aua ‘ten-pounder fish, Elops hawaiiensis’ < PPn *kafa ‘diamond-scaled mullet, Mugil vaigiensis’ (Rar ka’a ‘Mugil vaigiensis’, Samoan ‘afa ‘Mugil vaigiensis’, Tūvalu kahakaha ‘mullet’); Haw ‘ō(w)ā ‘split open’ < PCEPN *koofaa ‘split open’ (cf. Mao kōwhā ‘split open’).

14. *pofore ‘skinned’ is identified by Marck as an idiosyncratic Proto Marquesic assimilation, whereby PCEPN *pafore is reflected in Marquesic languages with po-rather than *pa-. However, it is simply a variant in the set Haw pahole, pohole, paholehole ‘bruised, skinned, scraped, peeled as cooked taro’. The Hawaiian set has a cognate in Tua pahorehore ‘peeled, scaled, scraped off’ and an ultimate source in the attachment of a prefix to PCP *fore ‘stripped, peeled’. There is thus no reason to think that this is a Marquesic innovation. While Mqa poho’e and Mva po’ore share with Hawaiian an assimilation in this case, the assimilation has two possible sources.

Sporadic assimilation of PEPN *a to *o before *f in a syllable with primary stress is distinctive of Hawaiian in the terms Haw ‘ōhi’a < PPN *kafika ‘Malay apple’ (cf. Mqa kehika, Mva ke’ika, Rar ka’ika); Haw kohana ‘naked’ < PNPN *tafa-ngā (cf. Mqa tahakahaka ‘clear, open’, Mva ta’a‘aga ‘bare’); Haw nahā, nohā

---

7 Compare also Haw mahole, maholehole, mohole ‘to bruise, skin, scrape, as a flesh wound, to injure the feelings’ with Tua mahore ‘be skinned, abraded, peeled, as by sunburn’ (Anaa).
< PEPn *gafaa ‘broken, shattered (of a hard object)’; Haw ‘ōhelo ‘a small shrub, Vaccinium reticulatum, with small, rounded, toothed leaves and red or yellow berries’ (cf. Mao tāwhero ‘a tree, Weinmannia racemosa, with smallish, toothed leaves, reddish when young’; see 2 regarding the irregular correspondence of Haw ‘with Mao tāwhero’).

While this assimilation is especially common in Hawaiian, it occurs throughout Central East Polynesian, e.g. PPN *taforaqa ‘whale’ > PNPn, PEPn *taforaqa > PCEPN *totofo(a) > Haw koholā, Mqa toho’aa, Mao tohoraa; PNPn *tongafiti ‘Tonga-Fiji’ > Haw konohiki ‘land overseer’, Mva togoiti8 ‘lord, prince’; PPN *tufunga ‘expert’ > PTah *tufunga ‘expert, priest’ > Mao tohunga, Tua tohuga; PPN *kawiki ‘ghost crab’ > PEPn *kaaviti > Mva kavitiviti ‘small species of crab’ > PTah *koofiti ‘ghost crab’ (via *kaafiti) Haw ‘ōhiki, Tah ‘ōhiti, Rar kō’iti.

Hawaiian also sometimes raises unstressed or secondarily stresses *a to *o after PEPn *f as in Haw ho’o-, ha’a- < PPN *faka- ‘causative prefix’ and Haw Hono-, Hana- ‘prefix to place names’ < PPN *fanga ‘bay’, e.g. Hono-lulu ‘peaceful bay’.

The other possible source of the assimilation in Hawaiian is assimilation of an antepenultimate a to o before penultimate o or u. Examples of assimilation before o include: PNPn *kanofi ‘flesh’ > PEPn *kanofi (mata) ‘eye, eye ball’ > Haw ‘ōnohi (cf. Mqa konohi mata ‘white of eye’); PNPn *katoa ‘all, whole’ > Haw ‘oko’a ‘different, whole’, (cf. Mqa kotoatoa ‘all’); Haw pāloke, pōloke ‘wobbly’; Haw palolo, pololo ‘talkative’; Haw mākoi, mōkoi ‘fishing pole’; Haw mākoihi, mōkoihi ‘to dig, a type of taro’; Haw mākole, mōkole ‘red eyed’; Haw ‘ōpuu < PCP *kapua ‘cloud’; Haw pahu-a, pohu-a ‘push’. Note also the rule reversal in Haw hopu-a, hapu-a ‘grabbed’ < PPN *hopu ‘grab’.

Some terms fit more than one condition for raising of *a to Haw o, e.g. Haw honua < PPN *fanua ‘land’; Haw ‘ōhua, ‘ōhua ‘juvenile form of certain fishes’ < *kaa + PPN *fuafua ‘young mullet or mullet-like fish’, Haw hōkū < PPN *fetuqu ‘star’, probably via *fatuu.

3.2 Marck’s Marquesic vowel changes

Marck (2000:98) lists six idiosyncratic vowel changes which occurred in Proto Marquesic: PCP *haere > PMq *here ‘go, walk’, PCEPN *muka > PMq *muko ‘growing tip’, PCEPN *taiti > PMq *teiti ‘child’, PCEPN *tua-ngaane > PMq *tu-ngaane ‘woman’s brother’, PCEPN *tokerau > PMq *tokolau ‘north’ (§3.1, 6), PCEPN *taokete > PMq *tokete ‘ego’s same-sex sibling-in-law’ (§3.1, 12).

Except for the change PCP *tokerau > PMq *tokolau, discussed in §3.1 under 6, each of these vowel changes also occurs in a language classified as Tahitic under the standard theory. Except for the term PCP *taiti, the vowel change occurs in at least one Tahitic language with the very lexical item listed by Marck. These six vowel changes are therefore only weak evidence that Hawaiian is a Marquesic language.

8 The expected glottal stop not indicated.
4 Terms uniquely shared by Hawaiian and Tuamotuan

Parallel to the seventeen Proto Marquesic innovations listed by Marck (1996) which remain valid after the fourteen in §3.1 are removed, the following list of 45 lexical innovations are unique to a hypothetical Proto Tuamotuic ancestral to Hawaiian and Tuamotuan. Tuamotuan terms listed below come exclusively from Stimson (1964) while the Hawaiian data are from Pukui and Elbert (1965). An attempt was made using dictionaries to verify that the unique terms collected do not also exist in Tahitian, Marquesan, Cook Islands Maori, or New Zealand Māori. POLLEX 1995 (i.e. the 1995 version of POLLEX) was also used to check for other cognates. A number of terms were eliminated and a closer look would certainly cast doubt on or eliminate more of these apparent uniquely shared innovations. Nonetheless, the total of 45 terms apparently uniquely shared by Tuamotuan and Hawaiian is comparable to the total of 31 apparent Proto Marquesic terms identified by Marck in POLLEX.

Besides listing the terms uniquely shared by Hawaiian and Tuamotuan, I also list possible derivations and information on Mangarevan when relevant. The information on Mangarevan will be discussed in more detail in §5.1.

1. Haw une, uneune ‘to pry’, Tua una ‘to pry, make a prying, prodding motion’ (Anaa). This cognate set is likely derived from PPn *quna ‘turtle carapace’ as flat pieces of turtle carapace are used for tools in Polynesian cultures. The Hawaiian form shows the same phonological change of $u(C)a > u(C)e$ found in §3.1 under 12 and §5.2 under 1.


3. Haw heiau, haiau ‘traditional religious structure parallel to a Tahitian marae’, Tua fare-heiau ‘the shrine of a maitu (spirit, god)’ (Vahitu), ‘assembly house’ (Tatakoto). The initial element *hei/-hai- of this word may be derived from an earlier PCEPn *hai/fai ‘to sacrifice’ > Haw hai, mōhai ‘sacrifice’, Tua hai ‘to place upon an altar in sacrifice to a god’ (Hao), Tah hāia ‘human sacrifice’.

4. Haw hi‘o-lani ‘sleep, lie at ease, lounge’, Tua hikohiko ‘intense drowsiness’, hiko ‘to be affected by some compelling force as drowsiness’ (Anaa). This term may be a metathesised version of PPn *foki > Haw ho‘i ‘return’ as in Hawaiian the term ho‘i is used to mean ‘go home’ with connotations of going home for the purpose of sleeping.


6. Haw kaka ‘to rinse off something’, Tua tata ‘clean, white, washed clean as sand unmixed with darker earth’ (Anaa).

7. Haw kali ‘vulva’, Tua tari ‘vulva’. Stimson related Tua tari to Tua tari ‘noose for catching birds, or eels’ (Vahitu). (Mangarevan has a cognate in Mva taritari ‘labia majora’.)

8. Haw kama ‘to tie on to’, Tua tama ‘to braid’ (Vahitu). This term may be derived from the initial two syllables of Haw kāma’a ‘braided sandal, shoe’, Rar tamaka ‘a covering or protection for the feet, a boot or shoe’.
9. Haw keiki ‘alu’alu ‘premature baby’ (cf. Haw keiki ‘baby, child’), Tua karukaru ‘newborn infant’ (Napuka). The narrow reference to a premature baby or newborn child is distinctive here. This term is related to terms with meanings having to do with loose skin, e.g. Tah ‘aru ‘loose skin’, Haw ‘alu’alu ‘loose, baggy, skin itself’, Tua karukaru ‘skin of living creatures’, Rar karukaru ‘hang or sag in folds, crumpled, wrinkled, wizened’.

10. Haw kō ‘fulfilled, come to pass’, Tua tō ‘to be fulfilled, come to pass’. This term may be related to PPn *tō ‘plant, conceive’ or to PPn *tō ‘to fall’.

11. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

12. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

13. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

14. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

15. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

16. Haw kū ‘fishhook with its point curved inward almost to the shaft’, Tua tū ‘to cut out, make, shape, fashion as a fishhook’ (Hao). This word is possibly related to PPn *tū ‘plant, conceive’ or to PPn *tū ‘to fall’.

17. Haw leleiona ‘remora fish, Echeneidae species’, Tua rerei ‘variety of fish, Gnathanodon species’ (Fangatau, Hao, Anaa, Amanu, Vahitahi, Hikueru). Mva rerei ‘to feel the stomach empty through hunger’ may be a cognate whose meaning changed through a metaphor relating to the attachment of the remora fish to the stomach of large fish.

18. Haw loea ‘an expert, especially in some art or sport’, Tua rōea ‘proficiency, expertness’, roe ‘be skilful, proficient, expert in a game or contest’ (Anaa).


20. Haw lole ‘to skin an animal’, Tua turore ‘to tear off, strip off skin’ (Fangatau). Mva rore (vowel quality uncertain) ‘to return or turn back when walking’ may be related, but does not share the same meaning.

21. Haw makamaka, kāmakamaka ‘fresh as fish’, Tua matamata ‘fresh of fish’ (Fangatau). This term is related to Haw maka ‘green as a fruit’, ‘raw as fish’ < PPn *mata ‘raw, unripe, of vegetables’, but there is no unreduplicated Tuamotuan form.
22. Haw *melumelu* ‘soft and slimy as spoiled raw fish’, Tua *meru* ‘to become soft’.

23. Haw *mu'o* *hala* ‘growing tip of pandanus root while still in air’, Tua *kāmuko* ‘the aerial root of the pandanus when it is barely fixed in the soil’ (Vahitahi). This term is related to Haw *mu'o* ‘leaf bud’, *mu'o* *hala* ‘growing young leaves of the pandanus’, Tua *muko* ‘the white heart, pith of the coconut tree’ (Anaa, Vahitahi, Takume), but its reference to a root is distinctive.

24. Haw *nai'a* ‘dolphin, porpoise’, Tua *gāika* ‘troop, band of animals, men, a flock of birds, a school, shoal of fish’ (Anaa). This term possibly derives from an earlier *ngā* *ika* ‘the-PL fish’.


26. Haw *pa'a* ‘reach the state of completeness as a house or lei when made, whole when speaking about eating taro, or other foods that might otherwise be processed in some way’, Tua *paka* ‘completely, entirely, all the way’ (Anaa). This term is likely related to Haw *pa'a* ‘firm, stuck, hard’, which in turn is likely related to earlier PEPn *paka* ‘dried, scorched, scab’. Rar VERB-*ā* *paka* ‘completely, to such a destructive or deplorable extent, a pejorative intensive’ may represent a historical step toward the meaning of completeness.

27. Haw *pikapika* ‘round suckers on tentacles of octopus’, Tua *pīpīta* ‘a circle’ (Ro), *pīta* ‘testicles’ (Vahitahi, Anaa), *pīta* ‘to make into a roll, as of dried pandanus leaves’ (Raeo, Anaa). Note that a Hawaiian *kūka'a* ‘circular, flat roll of pandanus leaves’ looks like a huge octopus sucker. The octopus connection suggests a possible relationship to Tah *pīta* ‘tempt as in using a shell to fish for octopus’.

28. Haw *wehi* ‘an ornament’, *wewehi* ‘lovely’, *ulu-wehi* ‘lush and beautiful’, Tua *vehi* ‘imposing, beautiful’ (Vahitahi). Tua *vehi* also means ‘awe-inspiring, terrible’. The existence of Mao *wehi* ‘be afraid, terrible’ suggests that the negative meaning is the older one with the innovative use of the word to indicate beauty deriving from a sense of awe similar to the experience of fear.


30. Haw *'ilio* ‘dog’, Tua *kurio* ‘dog’ (Ro, Taenga, Anaa, Hao, Amanu, Makemo, Fangatau). These terms derive from PPN *kulii* ‘dog’ with dissimilation of the last two vowels. Haw *'ilio* also shows assimilation and lengthening of the first vowel. Mva *kurio* (vowel quality uncertain) ‘the name of a large caterpillar’ may be related, but Mva *kuri* (vowel quality uncertain) ‘dog’ suggests otherwise.

31. Haw *'iwa* ‘frigate bird’, Tua *kiva* ‘male frigate bird’ (Anaa, Hao, Vahitahi). POLLEX 1995 suggests that Mao *Moana-Nui-A-Kiwa* ‘Pacific Ocean’ is related, but a connection to frigate birds in the Māori term is not clear. The term may be related through metathesis and glide formation to Rap *kuia* ‘booby bird’, Mao *kuia* ‘brown petrel, black petrel’, Tua *kuia* ‘a bird’.

32. Haw *'ōmu'o* ‘nipple-shaped end away from stem of coconut’, Tua *kōmuko* ‘the tip of the nipple, teat’ (Anaa). This term can be related to Haw *mu'o* ‘leaf bud, especially those that are twisted together and pointed’, Tua *muko* ‘the white heart,
pith of the coconut tree’ (Anaa, Vahitahi, Takume). Mva komuko (vowel quality uncertain) ‘a young coconut in which the flesh is still soft’ appears closely related.

33. Haw ‘ono ‘delicious, good tasting’, Tua kono ‘sweet, fresh as water, sugary, agreeable to the taste as food’ (Anaa). This term may be derived from PPN *kolo ‘intend or want to do something, desire’ > Tua koro ‘desire, hanker to do anything’, Haw ‘ono ‘crave’. The transformation into a stative verb along with the agreement in reflecting earlier *r with an *n indicates a shared history for the innovative Hawaiian and Tuamotuan terms. (See §3.1, 7 above on the problem of r/n substitution in East Polynesian.)

34. Haw ‘unu ‘twist the ankle’, Tua kūnu (Vahitahi), nuku (Anaa, Fangatau, Vahitahi) ‘to edge, hitch, slide the body along from a sitting position’, nuku ‘move the body slightly from its previous position’ (no dialect information). This correspondence is weakened by the fact that Tuamotuan reflects the non-metathesised form in the meaning most like the metathesised Hawaiian term. Hawaiian retains the non-metathesised form in Haw ‘a-nu’u ‘sprain, disjoined vertebra’ and this in turn is related to Rar ma-nuku ‘slip out, slip off, come undone, dislocated.’

35. Haw ‘āpē ‘mucus from the nose when one is crying’, Tua kupe ‘mucous excretion’ (Anaa, Fangatau). This term is related to Haw hūpē, Tua hupe ‘nasal discharge’ < PPN *isu-peqe ‘nasal mucus’.

36. Haw wa’ewa’e ‘cloven foot’, Tua vakevake ‘foot, leg’ (Napuka), cf. Haw wāwae ‘foot, leg’, Tua vaevae ‘foot, leg’ (no dialect information) < PPN *waqe ‘foot, leg’. Substitution of k and glottal stop for each other may have occurred sporadically in Proto Polynesian, as suggested by the doublet PPN *kawe ‘carry’, *qawe ‘take’ and Geraghty’s (1983:160–161) observation of the irregular correspondence PPN *q with dialectal Fijian k. Marck (2000:70) made a similar observation about PPN *h, *q > Mqa’. There are several sets of Hawaiian/Tuamotuan cognates that uniquely among East Polynesian languages reflect PEPN *q with reflexes of PEPN *k. These are listed below, in 37–40.

37. Haw maka-‘ala ‘observant, watchful’, Tua ma-kara ‘be alert, on one’s guard (Anaa) < PPN *mata ‘eye’ + *qara ‘awake’ (cf. Mao mataara ‘watch, keep awake’).

38. Haw kū-‘aha’aha ‘stand with hands on hips’, Tuaakahakaka ‘the crotch between the legs’ (Vahitahi) (cf. Rennellese ‘aha’aha ‘walk with legs apart’).


40. Haw māla’e ‘clear of weeds, as a field’, Tua mārakerake ‘be cleared open, bare as land’ < PPN *malaqe ‘open cleared space used as meeting-place or ceremonial place’ (cf. Tah marae ‘open place of pre-contact worship, cleared of wood, weed, rubbish, as a garden’).

There are also a number of unique Hawaiian retentions of PPN *q via *k, e.g. Haw lae-nihi, le’e-nihi ‘a type of fish with a square head (lit. ‘steep forehead’, cf. PPN *laqe ‘forehead’); Haw ‘eke‘eke cringe, shrink away from < PPN *faka-‘ete‘ete ‘act with caution’ > Mao whakaetti ‘loathe’. Tuamotuan includes a parallel retention, not found in Hawaiian, i.e. Tua reko ‘speak’, reo ‘voice’ < PPN *lego ‘voice’.

Two other examples of PPn *q being reflected irregularly as PEPn *s exclusively in Hawaiian and Tuamotuan cognates are given below in 42 and 43.

42. Haw *halo ‘belly fin, especially on a shark’, Tua *hāro ‘to make sweeping extended motion with the arms, flippers, wings’ (Anaa) < PPn *qalo ‘beckon’, ‘belly, bowls’, ‘to paddle’ (cf. Haw *alo-piko ‘belly of a fish’).

43. Haw *hono ‘mend a net’, Tua *hono ‘to join together end to end’ (Anaa) < PPn *qono ‘to mend a net’.

44. Haw *ma’ū ‘better than nothing, enough to satisfy given the circumstances’, Tua *makū ‘eat to satisfaction’ (no dialect information). This term is likely an irregular derivation from PPn *mahu ‘be abundant, plentiful of food’ < PEOc *mazu ‘sated’.

45. Haw *‘olu ‘cool, refreshing’, Tua *koru ‘cool, refreshing’ (Tapuhoe). Mva koru (vowel quality uncertain) ‘humid, very damp; said of lands sogged with water’ may be related, but is not semantically close.

5 Considering a realignment of East Polynesian subgroupings

5.1 Rejecting Tuamotuan as a fourth Marquesic language

In reconstructing an initial set of vocabulary for Proto Marquesic and another for Proto Tahitic, Green (1966) did not give any Tuamotuan data. However, citing Stimson’s (1964) introduction to the then newly published Tuamotuan dictionary, Green (1966:28) wrote, ‘Most of these dialects are in various ways related to Tahitian, but others are not and suggest origins at some point from the Marquesan subgroup.’ In a footnote, Marck (1996:510) stated that ‘the classification of Tuamotuan as a Tahitic language is misleading’ and expressed a feeling that Tuamotuan dialects ‘most notably those in the north, which may blend towards Marquesan, and those in the south, which may blend towards Mangarevan, Rapan, or others’ could be used by linguists to ‘demonstrate a more contiguous gradient of change between central Tuamotuan and Marquesan than the abrupt change that is implicit in most discussions of Eastern Polynesian language geography.’ Fischer (2001:119) too felt that Tuamotuan dialects were divided into two groups, but rather than claiming that the dialects geographically close to Mangarewa were Marquesic, he claimed that they shared with Mangarevan ‘vestiges’ of his Proto Southeast Polynesian.

Green’s statement that some Tuamotuan dialects have Marquesan origins is parallel to Stimson’s identification of the Tuamotuan spoken on the islands of Napuka and Reao, as well as part of Pukarua, as distinctive. Stimson felt that the unique features of the dialect of Napuka could be related to Marquesan and supported his speculation with local traditions that the first colonisers of Napuka came from the Marquesas. Stimson (1964:23) does state, however, that ‘... the aberrant forms, while apparently derived from Polynesian roots, do not indicate a close affiliation with Marquesan as spoken today.’ Of course, demonstration of a link to Marquesan must be with the Marquesan language as recorded, and claims that certain Napuka terms represented ‘old Marquesan’ could be countered by a proposal that
they were local Napuka developments, unless ‘old Marquesan’ was preserved in another Marquesic language, that is Hawaiian or Mangarevan.

The terms shared by ‘Marquesic’ Hawaiian and ‘Tahitic’ Tuamotuan listed in §4 above do not come primarily from any of the three Tuamotuan dialects Stimson identified as aberrant, that is Napuka in the north near the Marquesas, or Reao and Pukaruan in the south near Mangareva. Instead, the unique terms shared by Hawaiian and Tuamotuan are most commonly in items marked by Stimson as general Tuamotuan or as from the islands of Anaa and Vahitahi, which served as the major sources of data for Stimson. While Vahitahi is not far from Pukarua and Reao, Anaa is on the opposite side of the archipelago and is among the geographically closest Tuamotuan islands to the Society Islands. Thus, Hawaiian shares innovations with Tuamotuan dialects that are most commonly seen as ‘Tahitic’ within Green’s subgrouping proposal.

I have also been struck by the fact that the number of terms which I found to be shared uniquely by Hawaiian, Marquesan and Tuamotuan was small compared to the number shared exclusively by Hawaiian and Tuamotuan. Already discussed in §3.1 were the terms listed by Marck as Proto Marquesan *kaihue (Tuamotuan data suggest *kaihua), *kena (Tuamotuan data suggest *kena), and *ko(o)piko (Hawaiian and Tuamotuan data suggest *to(o)pio, *ko(o)pio.) These forms are all affected by late phonological rules, that is PEPn *r > Mqa’, k (distinctive of Marquesan); PEPn *tV(C)V(C)V(C)V > *kV(C)V(C)V(C)V (commonest in Hawaiian) and PEPn *u(C)a > PEPn *u(C)e (distinctive of Hawaiian, Tuamotuan and Tatitian.) The overlap of theses phonological rules in the various northern East Polynesian languages suggests borrowing.

The three terms listed immediately below are the only innovations found that show consonant phoneme correspondences consistent with regular reflexes of Proto East Polynesian and which are shared exclusively within East Polynesian by Tuamotuan, Hawaiian and Marquesan.

1. Haw hinihini ‘indistinct, faint as a voice, ‘ū-hini ‘grasshopper’, Tua hini ‘to whisper, speak with a low, muffled voice’ (Hao), hinihini ‘a variety of tiny cricket-like insect’ (Vahitahi), Mqa hinihini ‘cricket, its call’, Mva iniini (glottal stops not marked) ‘a locust, grasshopper’.

2. Haw houpo ‘diaphragm, chest, Tua hōupo ‘heart of a man or beast’ (Fangatau, Amanu, Hao, Vahitu), haupo ‘heart of a man’ (Fangatau), Mqa houpo ‘heart’, Mva oupo (glottal stops not marked) ‘the heart’.


As for a close relationship between Tuamotuan and Mangarevan, none of Marck’s proposed Proto Marquesan terms that I found to be shared by Tuamotuan and Hawaiian appear in Tregear’s Dictionary of Mangareva, but of the three cognates listed above as shared by Tuamotuan, Hawaiian, and Marquesan, two are shared by Manganevan. While Tregear’s Dictionary of Mangareva is certainly limited, I have been struck by the low number of cognates and near-cognates of the 45 terms listed in §4 as shared by Hawaiian and Tuamotuan but not by Marquesan and the Tahitic languages. Of the 45 Hawaiian/Tuamotuan terms, only one, 7, is phonologically and semantically fully cognate with a Mangarevan term. An additional six terms (16, 17, 20, 26, 30, 32) meet phonological criteria for cognacy but are semantically questionable. The 45 shared innovations of Hawaiian and Tuamotuan are thus not due to Tuamotuan being a Marquesic
language that has preserved Marquesic terms lost in Marquesan but preserved in Mangarevan. In summary then, Tuamotuan, Hawaiian and Marquesan uniquely share only six terms, some of which have phonological forms suggesting the possibility of borrowing.

Tuamotuan, Hawaiian and Mangarevan share seven terms, plus two more which overlap with those shared by Tuamotuan, Hawaiian and Marquesan. Lack of a substantial body of innovations uniquely shared by Tuamotuan, Hawaiian, Marquesan and Mangarevan (two terms), or even by Tuamotuan and Hawaiian and either Marquesan or Mangarevan (fourteen terms), is evidence against Tuamotuan being a member of the Marquesic subgroup as defined in the standard theory. Given the geographic proximity of the Marquesas and especially Mangareva to the Tuamotus, borrowing is to be expected among these languages. Thus few of the shared terms are likely to be due to inheritance from a uniquely shared ancestor. Fischer (2001:117) gives an example of an innovative term not found in Hawaiian and shared probably through borrowing by Marquesan, Mangarevan and Tuamotuan, but for example PEPn *kooqiro ‘type of eel’ > southern Mqa kō’e ‘fresh water eel’, Mva kōere ‘eel’, Tua kōere ‘a variety of fish’.

5.2 Reconsidering the Hawaiian–Tahitic relationship

If the 45 unique similarities between Hawaiian and Tuamotuan (and in a few cases Mangarevan) are not due to Tuamotuan being Marquesic, might they be due to Hawaiian being Tahitic? Under a hypothesis of Hawaiian as Tahitic, Hawaiian similarities to Marquesan would be due to borrowing and Hawaiian would subgroup among other languages with Tahitian, Māori, and Rarotongan. Hawaiian and Tuamotuan would then form a subgroup within Tahitic.

The following eleven innovative terms with regular correspondences among their consonant phonemes shared solely between Hawaiian, Tuamotuan and Tahitian suggest that a Tuamotuic subgroup of just Hawaiian and Tuamotuan would be part of a larger north-western East Polynesian subgroup within Tahitic containing Hawaiian, Tuamotuan and Tahitian. Data supporting such higher level subgrouping are given below with possible derivations for some terms.

1. Haw ānuenue ‘rainbow’, Tua anuenue, anuanua ‘rainbow’ (Vahitu, Vahitahi), Tah ānuenue, ānuanua ‘rainbow’ < PEPn *aunuanua ‘rainbow’. These reflexes follow the rule PEPn *u(C)a > u(C)e discussed in §3.1, 11 and §4, 1 and 11.

2. Haw manawa ‘time’, Tua manava ‘time’ (Fangatau), Tah manava hoi ‘early crop of breadfruit’. These terms are possibly related to PPn *manawa ‘breath, breathe’. Note that hoi in Tah manawa hoi may be related to Tua hoi ‘to be just off true, not quite at right angles’ (Hao), thus indicating a derivation of Tah manawa hoi as ‘time that is just off true for a crop of breadfruit’.

3. Haw pahi ‘knife, often of bamboo to peel taro’, Tua pahi ‘knife (obsolete)’ (Vahitahi), Tah pahi ‘piece of wood used to peel breadfruit’. This term is likely derived either from PPn *tapa-si ‘split, cut’ or PPn *fasi ‘split, cleave’.

4. Haw paoa ‘bad luck in fishing’, Tua pāoa ‘unlucky, unsuccessful’ (Vahitahi, Anaa), Tah pāoa ‘bad luck’. Note also Mva paoa (vowel length and glottal stops uncertain) ‘be poor; wandering; vagabond’.

5. Haw pāpio ‘young stage of various jacks called ulua’, Tua papio ‘a variety of caranax fish’ (Takaroa), Tah papio ‘a variety of fish’.
6. Haw kāmanomano ‘a variety of grass’, Tua tāmagomago ‘a variety of tall weed’ (Anaa), Tah tāma’oma’o ‘a species of grass’.

7. Haw kimo ‘a jack stone’, Tua timo ‘a small round worked stone used in throwing for distance’ (Anaa), timo-pei ‘a stone juggling sport’ (no dialect information), Tah timo ‘name of a play with stones’.

8. Haw kūmūmū ‘blunt’, Tua tūmū ‘blunt’ (Anaa), Tah tūmū ‘blunt’. This term possibly has a bimorphemic etymology cognate with Fijian mucu ‘blunt’, but with a distinctive prefix kū-.

9. Haw kīpē ‘to bribe’, Tua tīpē ‘to beg, go begging, as for food, favours, gifts’ (Vahitahih), Tah tīpē ‘beg, borrow’.

10. Haw kōkē ‘to knock together as ulu maika playing stones or billiard balls’, Tua tōtē ‘to strike, collide together’ (Vahitahi), Tah tōtē ‘knock against’. Note also Mva tote (vowel quality uncertain) ‘to strike a stone with a piece of wood’.

11. Haw ‘a’apo ‘quick in learning’, Tua kakapo ‘bright, intelligent’ (Anaa), Tah ‘a’apo ‘understand a thing quickly’. This term is derived from PPn *kapo ‘catch, grasp’.

Two of the terms in the above list, that is numbers 4 and 10, also have resemblant forms in Mangarevan, suggesting borrowing from one of the three languages into Mangarevan.

Connecting Hawaiian to the Tahitic subgroup beyond Tuamotuan and Tahitian are a number of innovations explained as borrowings in the standard subgrouping (Green 1966; Marck 1996, 2000). Green (1966:26) identified two of his five Proto Tahitian reconstructions as borrowed into Hawaiian, that is PTah *tufa, *tuha ‘spit’ > Haw kuha, Tua tuha, tufa, Tah tuha (cf. PPn *aqanu ‘spit, spittle’); PTah *waha > Haw waha, Tua, Tah vaha ‘mouth’ (cf. PPn *fafa ‘mouth’, Mqa haha, fafa, Mva ‘a’a ‘mouth’). Among Green’s other three Proto Tahitian terms, PTah *ki (more likely *kī) ‘full’, Tah ī (cf. PNPN *fofonu ‘full’) also has a likely Hawaiian cognate in Haw ī ‘tight, difficult to extract, as a plug’ (cf. Tua kī ‘full’, Tua kīkī ‘very full, full to capacity, to be filled, rounded out, plump’).

Marck (2000:117, 134, 230) listed additional features of Hawaiian as Tahitic, that is features posited as Proto Tahitian innovations and inherited from Proto Tahitian into its descendants. Some of these features are questionable due to the existence of cognates external to Tahitic and Hawaiian, that is Marck’s PTah *kimi ‘seek’ > Haw īmi (cf. Rap kimi ‘seek’); PTah *tokomauri ‘hiccup’ > Haw mauli-awa (cf. Tikopia tokomauri, Nuk leia togo oulu). In another instance, Hawaiian and Tahitian show an innovative form while Tuamotuan is conservative, that is Haw waele, wele, Tah vaere < PPn *wele ‘to weed’ (Tua vere), suggesting that the innovation postdates Proto Tahitic.

Uncontested Proto Tahitian reconstructions by Marck (2000:99–102, 115–117, 133–135) and noted by him as reflected in Hawaiian include Haw kahuna, Tah tahu’a, Rar ta’unga < PTah *tufuga ‘priest, skilled person’ < PPn, PEPn *tufunga ‘skilled person’ and four features that the Hawaiian kinship system shares with Tahitic (Marck 2000:230): (1) the innovative use of PTah *taane < PPn, PEPn *taqane ‘male’ as the masculine qualifier in Haw kupuna kāne, Tua tupuna tāne ‘grandfather’; (2) the innovative PTah *mata-hiapo ‘first born’ > Haw makahiapo, Tah matahiapo, Rar mata’ia; (3) the change PEPn *makupuna > PTah *mokopuna ‘grandchild’, Haw mo’opuna, Tua mokopuna, Tah moko-
tuā (Mva makupuna ‘grandchild’); (4) use of PPn *taqane ‘male’, Tah, Tua tāne, Haw kāne to also mean ‘husband’.

Other Hawaiian forms not identified by Marck (1996:505–509) in his list of 21 of 109 Proto Tahitic reconstructions in POLLEX are PTah *poo-reho ‘cowrie shell’ > Haw pōleholeho ‘small cowrie’, Tua, Tah pōreho ‘a variety of sea shell’ and PPn *mutie > PTah *matie ‘grass’ > Tah matie ‘a species of matted grass’, Tua matie ‘a small-bladed, self-spreading variety of grass’, Haw mānie-nie ‘various grasses similar to bermuda grass’.

Marck’s (2000:100–101) PPn *nonu > PTah *nono ‘Morinda citrifolia’ > Tah, Tua nono ‘Morinda citrifolia’ must be a post-Proto Tahitic innovation as Mangaiam retains nonu, and Manihiki shows nenu. Furthermore, Haw nono ‘red’ may be cognate. Morinda citrifolia is used to make a red dye in Hawai‘i. Marck’s (2000:222) identification of the shared innovation Haw makahiapo, Tah matahiapo ‘first born’ is not found in Tuamotuan and like nono, will not be counted here in our list of cognates shared by Hawaiian and Tahitian with Tuamotuan. It is likely that a careful comparison of Tahitian and Hawaiian will uncover a number of shared innovations restricted to those two languages, but such a study is beyond the scope of the present focus on terms shared by Hawaiian and Tuamotuan.

In summary then, Hawaiian shows a stronger pattern of shared innovations with Tuamotuan in combination with at least Tahitian (and sometimes also other Tahitic languages) than it does with Marquesic languages. Post-PCEPn innovations shared by Hawaiian, Tuamotuan and Tahitian at present include 20 terms, that is three general Tahitic terms from Green (1966), four more identified by Marck (2000), and two missed by Marck (1996), plus the eleven additional uniquely shared innovations identified above.

Further evidence that Hawaiian is not Marquesic is the wide distribution in Hawaiian of the phonological change PPn *f > *w (and *p) noted by Green (1966:18, 24, 26) and Fischer (2001:116) as a later change in East Polynesian, and in some terms, dividing Marquesic from Tahitic, e.g. PPn *fafa ‘mouth’, Mqa haka, fafa, Mva ‘a’a, Haw waha, Tah vaha. As shown in §3.1.13, Hawaiian exhibits this change in a number of forms where it does not occur in Marquesic, or even, in some cases, in Tahitic.

The 20 innovative terms shared by Hawaiian, Tuamotuan and Tahitian are considerably more than the six terms (three of which show evidence of irregular consonant correspondences) shared by Hawaiian, Tuamotuan and Marquesan. Overall the 20 innovations shared by Hawaiian, Tuamotuan and Tahitian, taken together with the 45 innovations restricted to Hawaiian and Tuamotuan, give a total of 65 innovations defining a relationship between Hawaiian and Tuamotuan within a larger Hawaiian-Tuamotuan-Tahitian relationship. This number is much greater than the seventeen uncontested innovations shared solely by Hawaiian and Marquesan (or Hawaiian, Marquesan and/or Mangarevan) taken from Marck (1996:500–505).

A problem for a scenario in which Hawaiian is basically Tahitic is to explain the large number of terms shared exclusively by Tuamotuan and Hawaiian and the even larger number of widespread Tahitic terms found in Tuamotuan, but not found in Hawaiian. Marck (1998:505–508) noted that there were 109 Proto Tahitic reconstructions as against 29 Proto Marquesic reconstructions in POLLEX. He provided a sample of 21 of those Proto Tahitic reconstructions. Only three of these have Hawaiian cognates, while 17 (including the three with Hawaiian cognates) have Tuamotuan cognates. Reflecting such a small portion of the set of innovations defining the Tahitic subgroup, Hawaiian cannot be considered fully Tahitic in the sense that Tuamotuan can.
How is it that Hawaiian tends to share unique Tuamotuan words while avoiding more widespread Tahitic words? The common assumption of linguists and anthropologists interested in the settlement of Polynesia has been that populations moved to Hawai‘i, rather than considering that there were also movements from Hawai‘i to other areas. Borrowing of considerable numbers of terms from Hawaiian into Tuamotuan could explain how Hawaiian has failed to include many widespread Tahitic terms while still maintaining many shared innovations with Tuamotuan. Positing borrowing from Hawaiian into Tuamotuan, and also into Marquesan and Tahitian, does not exclude Hawaiian borrowing from these languages as well.

6 Hawai‘i’s ‘founder’ and ‘intrusive’ populations

Green’s (1966:26–27, 29–30) proposal that Hawaiian is a Marquesic language with Tahitic borrowings is part of a larger hypothesis regarding the prehistory of Hawai‘i shared with anthropologists who propose a secondary population intrusion from the Society Islands occurring several centuries after initial settlement from the Marquesas (Kirch 2000:245, 290–293). This hypothesis is challenged first by the much greater size of the Tahitic element in Hawaiian than of its Marquesic element. Unlike several other East Polynesian languages, which are spoken on small islands, Hawaiian occupies a large land mass that anciently supported a large population. It is difficult to see how a Tahitian element could overwhelm an indigenous Marquesic element in Hawai‘i once a significant population had developed. If somehow the language of Tahitian immigrants to Hawai‘i had overwhelmed a distinct indigenous Marquesic language, one could expect the remaining Marquesic elements to form a semantically marked substratum, as with the Rennelense ‘hiti’ substrata distinguished by a preponderance of terms for inland flora and fauna (Elbert 1988:266–277, 279–283). There is no such semantic unity within the Marquesic element of Hawaiian.

One could argue the opposite of Green’s proposal, namely that Hawaiian was originally Tahitic with a smaller, but still significant Marquesan element added by an intrusive population from the Marquesas. In support of such a proposal we have in §3.1 under 8 and 9 evidence that Hawaiian probably borrowed from Marquesan at a later date after the consonant inventories of Hawaiian and Marquesan diverged from each other. However, there is also phonological evidence of late borrowing in Hawaiian from Tahitian, e.g. Haw hi‘o-na ‘appearance’, Tah hi‘o, Tua higo ‘look’ < PTah *hingo ‘look’, where the expected directly inherited form in Hawaiian would have an n rather than a glottal.

While there is evidence of late borrowing from Marquesan and Tahitian, neither the phonologically marked Marquesic elements nor the phonologically marked Tahitic elements of Hawaiian have semantic distributions typical of a superstratum where a dominating intrusive population overcomes an indigenous group. Rather than having either a founding Marquesic or Tahitic population followed much later by an intrusive population from one of those two groups, Hawai‘i may have been settled at an early period before there were either distinctive Marquesic or Tahitic subgroups. Contact probably continued between Hawai‘i and other northern East Polynesian inhabited islands from the time of initial settlement, when there was essentially one language spoken in all of northern East Polynesia, until a considerably later period after which the languages of the various northern East Polynesian island groups had diverged and were marked by distinct consonantal phonologies. While drawing the line between an ‘intrusive population’ and ‘normal borrowing’ might be difficult in the case of Hawaiian, the linguistic evidence of
scattered words without highly distinctive semantic divisions suggests highly integrative contact with those who came to Hawai‘i from other island groups.

Evidence for phonologically marked late borrowing into Hawaiian and Tuamotuan from Marquesan was noted under §3.1. There is also phonological evidence for late borrowing from Hawaiian into Tuamotuan, e.g. Tua *makuahīne ‘mother, aunt’, cf. Haw *makuahine ‘mother, aunt’ (< PEPn *matuqa ‘parent’ + *wahine ‘woman’) and Tua *hānu ‘to hunt, serve for by means of the sense of smell’ (Anaa) (cf. Haw *hanu, ‘breathe’, *hanuhanu ‘sniff around for something’ < PEPn *fagu ‘breathe’, Tua *hegu ‘to snort, as a pig, horse’).

7 Innovations with other distributions

There are overlapping innovative isoglosses connecting Hawaiian with various subsets of East Polynesian languages that reach over the boundaries of the current Tahitic and Marquesic subgroups, and even beyond to include Rapanui. To the isoglosses already given, linking the languages of geographic northern East Polynesia, we can add the following two sets uniquely shared by Hawaiian, Tahitian, Tuamotuan and Marquesan: (1) Haw *lo’a, Tua *rōka, Tah *roa’a, Mqa *oa’a, koana, koaka ‘be obtained, be found’ (cf. Rar *rauka ‘get’, Mao *rawaka ‘abundant, sufficient’ and possibly Mva *rauga ‘said of a man to whom an object belongs’) and (2) Haw *malihini, Tua *arihini, manihini, Mqa *manihiti, ma’ithi, Tah manihini, manihiri, manuhini ‘guest, visitor’ (cf. Mao manuhiri, Rar *manu’iri and possibly Mva *manu’iri ‘to play at blindman’s bluff’). A third is Haw kā ‘oe ‘your, yours’ (found only in a traditional saying where it replaces normal kāu), Tua tā koe (part of a series where koe replaces PPN *u in possessives) (no dialect information), Mqa tā koe, tā ‘oe (part of a series where koe, ‘oe replaces PPN *u in possessives), Tah tā ‘oe (part of a series where ‘oe is the normal representative of second person singular in possessives although -u is used in rare variants), Mva to koe ‘your’. Also marking off the northern East Polynesian languages is the innovative use of reflexes of PPN *ko and/or zero as nominative marker with proper names found in Hawaiian, Tahitian, Tuamotuan and Marquesan rather than the reflexes of PPN *qa ‘personal article’ found in Easter Island, Rarotongan and New Zealand Māori. Further connecting the four languages are what may be identical irregular developments of earlier PPN *saa-ware ‘saliva, spit, slime’ and *wale ‘viscous fluid’ found in Haw hā ae ‘saliva’, Tua hākae ‘saliva’ (G), Tah hā‘ae ‘saliva’, Mqa ka’e ‘saliva’, Tua kae ‘saliva’ (Reao), *spittle (Vahitahi) and Haw *ae ‘liquid that oozes out of something such as prepared seaweed’.

There are also isoglosses that link Hawaiian, Tahitian, Marquesan, Tuamotuan and Rarotongan, for example the replacement of the irrealis predicate possession markers PPN *maqa, *moqo, maintained in Māori, Penrhyn, and Mangareva (Clark 1976:112–113), with realis predicate possession markers PPN *na(q)a, *no(q)o in Hawaiian, Tahitian, Tuamotuan, Marquesan, and Rarotonga. Linking a subset of these languages is the term for ‘parent-in-law’: Rar ʻongovai, Tah ho’ovai, and Haw makua hōnō(w)ai (doublet with makua hūnō(w)ai) < PCEPN *fungawai ‘parent-in-law’ > Mao hungawai; the particle of doubt Haw, Tua (Anaa), Tah, paha, Rar pa’a < PPN *peafa > Rap peaha, Mao pea; and the cognate set Rar, Tua tēia, Tah tēie, and Haw kēia ‘this’ and replacing PPNs *teene ‘this’ which is retained in Māori and Marquesan. Many other innovations are confined solely to a core East Polynesian area consisting of Tahitian, Tuamotuan, and Rarotongan, for example, Tah hōro’a, Tua horōga, Rar ʻōronga ‘to give’.
As noted by Langdon and Tryon (1983:42–43) Rapanui shares a number of terms with Marquesan and/or Mangarevan. Among those that appear to be innovative and are also shared with Hawaiian (and in one case Tuamotuan) are Haw *lauoho, Rap *rouoho, Mva *rouoho, Mqa *auoho ‘(head) hair’, (cf. Tah *rouru < PPN *lauqulu ‘hair of head’ > Haw *loulu ‘fan palm’, Mva *rouru ‘the completely dressed head of hair with all ornaments’); Haw hā-mau ‘silent’, Rap *mou ‘be silent’, Mqa *mou ‘peace (possibly related to PPN *malu ‘protect’ > Haw *malu ‘protection’, *malu-hia ‘peace’), and Haw *uhane, Tua *uhane, Mqa, Tua, Mva, Rap *kuhane ‘spirit’. In other cases, Rapanui, Marquesan and Mangarevan share an innovation and Hawaiian retains a Proto Polynesian form, e.g. PPN *lango ‘fly species’ > Haw nalo, Rap, Mva takaure ‘fly’, Mqa *tikau’e ‘fly’.

8 A wave model for northern East Polynesian

The many overlapping isoglosses linking Hawaiian with various languages within East Polynesian do not fit the standard tree model used by Green (1966) in first classifying Hawaiian as a Marquesic language later heavily influenced by an intrusive Tahitian population. Much more suitable is Pawley’s (1981) model for the development of the different branches of Oceanic based on rapid settlement followed by continued contact within the settled area that resulted in localised centers of innovation. When applied to northern East Polynesia Pawley’s proposal implies sustained maintenance of a dialect chain for a considerable time period over the largest expanse of open ocean regularly crossed by human beings using a neolithic technology.

Whatever the East Polynesian source of the first speakers of Pre-Hawaiian in Hawai‘i, this language likely differed little from Proto Central East Polynesian, or even possibly Proto East Polynesian.

Phonological evidence for a very early East Polynesian language being spoken in Hawai‘i can be found in the terms in §3.1, 13 where Hawaiian has the unique reflex Haw pu- < PPN *fu-, where all other East Polynesian languages merge PPN *fu- and *su- as *hu-, and the terms in §4, 40 where Hawaiian uniquely among Central East Polynesian languages has a consonantal reflex of PPN *q rather than the zero.

The grammar and morphology of Hawaiian is less conservative than that of Rapanui or even New Zealand Māori, but, there are a number of unique lexical retentions found only in Hawaiian among East Polynesian languages, e.g. Haw huā ‘envy’, jealousy’ < PPN *fuqa ‘jealous’ (cf. PCT *sae ‘jealous’); Haw pola ‘banana flower bud’ < PPN *polata ‘banana stem’; Haw kauhuhu ‘ridgepole’ < PPN taqufu (cf. PCT *taaflu ‘ridgepole’ > Mao tāhuhu, Mqa tahuhu); Haw kela, kelakela ‘excel, high’ < PPN *telua upper, frontal surface’; Haw kepa ‘look sideways, cut obliquely’ < PPN *tepa ‘look sideways, cut obliquely’; Haw kukuma, ‘ala-kuma ‘types of squat crabs’ < PPN *tuma ‘body louse’, Haw loha ‘variant name for the kilu, a game played with discs’ < PPN *lafo ‘a throwing game played with discs’; Haw mā’ila ‘clear, as the sea on a sunny day when the depths can be seen’ < PPN *ma(a)-kila ‘to appear under special conditions’; and possibly Haw hūkā-kai ‘brackish water’ < PPN *tufu ‘fresh-water spring, usually on the beach’. Hawaiian also retains some terms otherwise only found in Rapanui among East Polynesian languages, e.g. Rap hahave ‘flying fish’, Haw hahape ‘milk fish’; Rap keo ‘stomach’, Haw kā-‘eo ‘full, of a calabash, knowledgeable (stomach is the seat of knowledge)’; Rap nanai, Haw nananana ‘spider’; Rap kokope ‘club footed’, Haw o’opa ‘crippled,lame’; Rap mangai ‘fish hook’, Haw Mānai-a-kā-lani ‘famous fish hook of the demigod Māui’, māna ‘lei needle’ (cf. and also possibly Rap avai, Haw hā-‘awi ‘give’. As noted by Marck (2000:99)
A Tuamotuan challenge to the subgrouping of Hawaiian

311

and Langdon and Tryon (1983:42) Hawaiian also shares some retentions with Tahitic where Marquesic and Rapanui innovate, e.g. Haw malo < PPn *malo ‘loin cloth’ (but PMq *hami, Rap hamni) and other retentions with Marquesic where Tahitic innovate, e.g. Haw uluhe < PPn *hulufe ‘type of fern’ (but PTah *aruhe).

Whatever the immediate source of the initial East Polynesian settlement of Hawai‘i, the history of Hawaiian clearly involved a long period of extensive contact, common development, and mutual influence shared with other languages of East Polynesia, especially those of the Tuamotu, Society, and Marquesas Islands. There were also probably East Polynesian languages spoken on the long abandoned Line and Phoenix Islands lying between Hawaiian and the rest of northern East Polynesia, which also had contact with Hawaiian and other northern East Polynesian languages.

Based on the relative number of uniquely shared innovations, the Hawaiian portion of the early dialect chain had much more contact with its Tuamotuan portion than with its Marquesan portion. Additionally, a considerable number of lexical innovations characteristic of both Tuamotuan and Tahitian are found in Hawaiian. Within the northern East Polynesian dialect chain the Tuamotu and Society Islands likely formed a relatively uniform centre for a considerable period. Internal interaction within that centre, combined with considerable Hawaiian contact with the various components of that centre, would explain the pattern of shared innovations that includes just Hawaiian, Tuamotuan and Tahitian.

The various languages of northern East Polynesia outside Hawaiian have also had a long history of contact with each other and with other surrounding East Polynesian languages in the Cooks and Australs. Contact, especially between the Tuamotu/Society Islands centre and the nearby languages of the northern and southern Cooks and the Australs, resulted in many other innovations spreading within the central area of East Polynesia, only some of which were shared with Hawaiian.

As shown in §3.1 under 5 and 6, borrowing is also a likely source of some of the terms shared by certain northern Outlier languages and certain northern East Polynesian languages. Wilson (1985, 2008) provided evidence for a low-order genetic link between the languages of these two areas. Within Pawley’s wave model, borrowing between the northern Outliers and East Polynesian would be consistent with the initial settlement of East Polynesia from the northern Outlier region followed by diminishing contact over time. Some of this contact appears to have occurred simultaneously with the contact that was occurring internal to East Polynesian. Thus some of what appear to be unique Hawaiian retentions from Proto Polynesian may be the resture of borrowing from the Northern Outliers.

The pattern of overlapping isoglosses within East Polynesian is not inconsistent with Hawaiian tradition that names a wide variety of places with which Hawaiian contact existed in ancient times. Among the names of such places are Kuaihelani, Nu‘umealani, Melemele, and Polapola. The over-arching name, however, for all these places, was Kahiki, a name cognate with Tahiti, the name of the largest island outside of Hawai‘i in northern East Polynesia. The use of a single island name to represent distant lands beyond one’s home is characteristic of the Central East Polynesian languages. Cognates of Haw Kahiki are sometimes used in this way outside Hawai‘i, for example Mqa Tehiti ‘foreign land’, Mao Tawhiti ‘a distant traditional homeland’, Tua Tahiti ‘remote place’ (< PCEPn *Tafiti). Much more common outside Hawai‘i, however, is a version of PCEPn *Hawaiki, a name cognate with Hawai‘i, the largest island in the Hawaiian chain.
It is intriguing to consider that the name *Hawaiki might represent the southern counterpart of the Hawaiian term Kahiki denoting external lands, and that those two names anchored a long period of north–south contact within the northern East Polynesian linguistic area, whose memory is preserved in the traditions of the languages of Central East Polynesia referring to mythical places cognate with *Hawaiki and *Tafiti.

### Appendix 1: Languages and language abbreviations

#### Protolanguages

<table>
<thead>
<tr>
<th>Proto Central East Polynesian [PCEPn]</th>
<th>Proto Marquesic [PMq]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto Central Pacific [PCP]</td>
<td>Proto Nuclear Polynesian [PNPn]</td>
</tr>
<tr>
<td>Proto Eastern Oceanic [PEOc]</td>
<td>Proto Polynesian [PPn]</td>
</tr>
<tr>
<td>Proto East Polynesian [PEPn]</td>
<td>Proto Tahitic [PTah]</td>
</tr>
</tbody>
</table>

#### Tongic languages

| Niuean | Tongan |

#### Southern Outliers and Triangle Nuclear Polynesian languages

<table>
<thead>
<tr>
<th>East Futunan</th>
<th>Samoan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kapingamarangi</td>
<td>Takuu</td>
</tr>
<tr>
<td>Luangiua (Ontong Java)</td>
<td>Tikopia</td>
</tr>
<tr>
<td>Nukuoro</td>
<td>Tūvalu (Ellice)</td>
</tr>
<tr>
<td>Rennell</td>
<td></td>
</tr>
</tbody>
</table>

#### East Polynesian languages

<table>
<thead>
<tr>
<th>Hawaiian [Haw]</th>
<th>Penrhyn</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Zealand Māori [Mao]</td>
<td>Rapanui (Easter island) [Rap]</td>
</tr>
<tr>
<td>Mangarevan [Mva]</td>
<td>Rarotongan [Rar]</td>
</tr>
<tr>
<td>Marquesan [Mqa]</td>
<td>Tahitian [Tah]</td>
</tr>
<tr>
<td>Mōriori</td>
<td>Tuamotuan [Tua]</td>
</tr>
</tbody>
</table>
Appendix 2

Regular consonant correspondences between Hawaiian, Tuamotuan and all other languages for which examples appear in this paper are given in Table 1.

### Table 1: Regular orthographic reflexes of Proto Polynesian consonants

<table>
<thead>
<tr>
<th>Language</th>
<th>*p</th>
<th>*t</th>
<th>*k</th>
<th>*q</th>
<th>*f</th>
<th>*s</th>
<th>*h</th>
<th>*m</th>
<th>*n</th>
<th>*ng</th>
<th>*w</th>
<th>*r</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPN</td>
<td>*p</td>
<td>*t</td>
<td>*k</td>
<td>*q</td>
<td>*f</td>
<td>*s</td>
<td>*h</td>
<td>*m</td>
<td>*n</td>
<td>*ng</td>
<td>*w</td>
<td>*r</td>
</tr>
<tr>
<td>PTongic</td>
<td>*p</td>
<td>*t</td>
<td>*k</td>
<td>*q</td>
<td>*f</td>
<td>*h</td>
<td>*m</td>
<td>*n</td>
<td>*ng</td>
<td>*w</td>
<td>*r</td>
<td>*Ø</td>
</tr>
<tr>
<td>Tongan</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>Ø</td>
</tr>
<tr>
<td>MPn</td>
<td>*p</td>
<td>*t</td>
<td>*k</td>
<td>*q</td>
<td>*f</td>
<td>*h</td>
<td>*m</td>
<td>*n</td>
<td>*ng</td>
<td>*w</td>
<td>*r</td>
<td>*Ø</td>
</tr>
<tr>
<td>Tuvalu</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>E Futuna</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>K’marangi a</td>
<td>b/p</td>
<td>d/t</td>
<td>g/k</td>
<td></td>
<td>h</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>w</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Nukuoro b</td>
<td>b/p</td>
<td>d/t</td>
<td>g/k</td>
<td></td>
<td>h</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Luangiua</td>
<td>p</td>
<td>k</td>
<td></td>
<td></td>
<td>h</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Remnelse</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Samoan</td>
<td>p</td>
<td>t</td>
<td></td>
<td></td>
<td>f</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>g</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Nukulae</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>g</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Takuu</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>v</td>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Tikopia</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>f</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>v</td>
<td>l</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>PEPn</td>
<td>*p</td>
<td>*t</td>
<td>*k</td>
<td>*q</td>
<td>*f</td>
<td>*s</td>
<td>*Ø</td>
<td>*m</td>
<td>*n</td>
<td>*ng</td>
<td>*w</td>
<td>*r</td>
</tr>
<tr>
<td>Rapanui</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>PCEPn</td>
<td>*p</td>
<td>*t</td>
<td>*k</td>
<td>*Ø</td>
<td>*f</td>
<td>*s</td>
<td>*Ø</td>
<td>*m</td>
<td>*n</td>
<td>*ng</td>
<td>*w</td>
<td>*r</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>p</td>
<td>k</td>
<td></td>
<td></td>
<td>h/p</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mori or</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>wh</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>w</td>
<td>r</td>
<td>r</td>
</tr>
<tr>
<td>N Marquesan</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>k</td>
<td>v</td>
<td>l/r</td>
<td>l/r</td>
</tr>
<tr>
<td>S Marquesan</td>
<td>p</td>
<td>t</td>
<td>/k</td>
<td></td>
<td>h/p</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mangarevan</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penrhyn</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>s</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Rarotongan</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h</td>
<td>h</td>
<td>m</td>
<td>n</td>
<td>ng</td>
<td>v</td>
<td>l</td>
<td>l</td>
</tr>
<tr>
<td>Tahiti</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td>h</td>
<td></td>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuamotuan</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td></td>
<td>h/p</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

a The phonetic realisation of Kapingamarangi and Nukuoro orthography is b = IPA /p/, p = IPA /pp/, d = IPA /t/, t = IPA /tt/, g = IPA /k/, and k = IPA /kk/.

b The representation of Marquesan here as separated into Northern and Southern dialects ignores the issue of dialect mixing that has likely occurred in the history of Marquesan language varieties.

### References


(Abbreviated POLLEX in the text.)
William H. Wilson


~ Grammar and use ~
Numerals in Rongga

I WAYAN ARKA

1 Introduction

Rongga is a small Austronesian language spoken by around four thousand people on Flores Island, Indonesia. It has been increasingly marginalised, under pressure not only from the national language (Indonesian) but also from its more dominant neighbours, particularly Manggarai and Waerana (see Arka 2005; 2008 for details). While Rongga is still used in everyday life and acquired by children, particularly in the interior areas of its territories, its domains of use are rapidly shrinking. Among the diminishing domains are ritual language (Arka 2009) and certain everyday expressions, such as numerals.

It has been pointed out by Comrie (2005) that numeral systems are particularly susceptible to sociolinguistic changes (through language contact) that, over time, results in their extinction. This is the case with Rongga’s numerals. This paper reports and discusses the findings of research on numerals in Rongga as part of a larger language documentation project on Rongga. It will be shown that there are two quite different numeral systems in Rongga. For convenience, the systems are referred to as the restricted system and the regular system, respectively. The restricted system is already extinct, and the other is highly endangered.

The paper is outlined as follows. The numerals of the two systems are presented in §2. Then, their descriptions are given in §3. These include their contemporary functions (§3.1), their structural properties in relation to classifiers (§3.2), and their formal and arithmetical systems (§3.3–§3.4). Finally, a brief discussion of the Austronesian connection, language contact and endangerment is presented in §4.

1 Part of this paper was already presented at the MLI (Masyarakat Linguistik Indonesia, or Indonesian Linguistic Society) conference in Padang, Indonesia (July 2005) and ICAL (International Conference on Austronesian Linguistics, Palauan, January 2006). Research on Rongga was made possible by the ELDP and NSF grants (IPF0011, BCS-0617198). The numeral data in Manggarai, Ngadha, and Ende in the appendix were given by Jeladu Kosmas, Watu Yohanes Vianey and Frans Remba, all are native speakers of the Manggarai, Ngadha, and Endenese languages, respectively. I thank Margaret Florey and John Bowden, for comments on the earlier version of this paper, and Malcolm Ross for his excellent last-minute comments and suggested changes to improve the presentation of the paper.
2 The numerals

The restricted and regular Rongga numeral sets are given in (1)–(2) below. For ease of exposition, they are represented in two parts, one to ten in (a) and eleven to twenty in (b). (Higher numerals are given in Appendix A.)

(1) a. Restricted:2

\[
\begin{array}{llll}
\text{dha} & \text{\textquoteleft one\textquoteright} & \text{woe} & \text{\textquoteleft six\textquoteright} \\
\text{dhua} & \text{\textquoteleft two\textquoteright} & \text{sipi} & \text{\textquoteleft seven\textquoteright} \\
\text{demu} & \text{\textquoteleft three\textquoteright} & \text{sapa} & \text{\textquoteleft eight\textquoteright} \\
\text{dheke} & \text{\textquoteleft four\textquoteright} & \text{nggwo} & \text{\textquoteleft nine\textquoteright} \\
\text{ali} & \text{\textquoteleft five\textquoteright} & \text{ngguru} & \text{\textquoteleft ten\textquoteright} \\
\end{array}
\]

b. \text{esanggesa} ‘eleven’ \text{anggunae} ‘sixteen’
\text{zhuandua} ‘twelve’ \text{nenggonae} ‘seventeen’
\text{telunggetu} ‘thirteen’ \text{soroila} ‘eighteen’
\text{wutunggutu} ‘fourteen’ \text{watupesa} ‘nineteen’
\text{limakima} ‘fifteen’ \text{mopla} ‘twenty’

(2) a. Regular:

\[
\begin{array}{llll}
\text{(esa)} & \text{\textquoteleft one\textquoteright} & \text{limaes} & \text{\textquoteleft six\textquoteright} \\
\text{zhua} & \text{\textquoteleft two\textquoteright} & \text{limazhua} & \text{\textquoteleft seven\textquoteright} \\
\text{telu} & \text{\textquoteleft three\textquoteright} & \text{zhuambu} & \text{\textquoteleft eight\textquoteright} \\
\text{wu} & \text{\textquoteleft four\textquoteright} & \text{taraesa} & \text{\textquoteleft nine\textquoteright} \\
\text{lima} & \text{\textquoteleft five\textquoteright} & \text{sambulu} & \text{\textquoteleft ten\textquoteright} \\
\end{array}
\]

b. \text{sambulu saesa} ‘eleven’ \text{sambulu esa limaes} ‘sixteen’
\text{sambulu esahzza} ‘twelve’ \text{sambulu esa lima} ‘seventeen’
\text{sambulu esatelu} ‘thirteen’ \text{sambulu esa zhuambu} ‘eighteen’
\text{sambulu esawat} ‘fourteen’ \text{sambulu esa taraesa} ‘nineteen’
\text{sambulu esalima} ‘fifteen’ \text{mbulu zhu} ‘twenty’

In what follows, the functions and internal systems of the two sets are compared. To understand the systems, it is first necessary to discuss the forms and structures of the numerals in Rongga in relation to the system of classifiers in this language.

3 Functions and forms

3.1 Contemporary use

The numeral set exemplified in (1) is dubbed ‘restricted’ because it is known to have a restricted function. It has been suggested that the numerals of this set were used in a language game in the old days. There may be other functions of this set, but the community elder (Mr Antonius Gelang) consulted for this study had only a very vague recollection of the counting in a game using this set. He also had problems in remembering other
functions of the numeral set. Language games with strange or irregular (counting) patterns are also reported in other Indonesian languages of Maluku.\textsuperscript{3}

The regular set, in contrast, is used for counting in everyday life. However, it now competes with the Indonesian numeral set, which is increasingly used by Rongga speakers, especially children. This leads to an endangerment of the traditional counting set, further discussed in §4.2.2.

### 3.2 Numerals and classifiers

Numerals express quantities of entities. Entities are sorted by classifiers in Rongga. A numeral, therefore, almost always appears with a sortal or mensural classifier in Rongga, forming a numeral phrase (NumP).\textsuperscript{4}

At the broadest level, sortal classifiers in Rongga encapsulate the natural classifications of human, animal, and plant. The category of human gets different classifiers from that of animal, which in turn receives different classifiers from that of plant.\textsuperscript{5} The following are common sortal classifiers found in Rongga:

(3) i. Classifiers for living things:
   a. human: \textit{ata} or \textit{mori}
   b. animal: \textit{eko}
   c. plant: \textit{pu’u}

ii. Classifiers for non-living things: specific shape-, part- and property-based
   a. fruit related (round(ish) and cube-like objects): \textit{li’e}
   b. leaf related (flat, finger-like shape): \textit{wunu, kambi}
   c. log related (round and long): \textit{toko}
   d. wood related (flat, thick, and long, not easily bent): \textit{mbi’i}
   e. wood related (flat and thin, easily bent/folded): \textit{nolo}
   f. dead trees lying on the ground: \textit{fata}

iii. Default classifier for items not easily classified in (i) and (ii), typically for inanimate things: \textit{esa}

Classified with living things in (3) are those that can grow and die. Non-living things are sorted in terms of their different attributes, such as size and texture, as shown in (3). Hence, a living, standing tree is referred to using \textit{pu’u}, while a dead tree lying on the ground is referred to using \textit{fata}, not \textit{pu’u}.

Structurally, a numeral phrase (NumP) is part of a noun phrase (NP) which can consist of other units such as a possessive pronoun and/or an adjunct. The NumP may appear in either of two positions, before or after the noun head. Thus, both structures in (4) are equally acceptable for ‘a chicken’ in Rongga.

---

\textsuperscript{3} Thanks to Margaret Florey who pointed this out to me.
\textsuperscript{4} NumP containing classifications exemplified in this paper are encountered in regular numerals only. No restricted numerals are attested in the author’s Rongga corpus. It remains unclear whether the restricted numerals could appear with a classifier.
\textsuperscript{5} God or spirits could be classified as human, taking \textit{mori}. However, native speakers of Rongga are often uneasy about assigning \textit{mori} to God or the spirits and say they are not things of which we know their number. The use of \textit{mori} is certainly better than \textit{ata} for God(s), for example \textit{samori Dewa} ‘a/the God’ is often considered acceptable but \textit{sangata Dewa} ‘the/a God’ is not.
(4) a. [manu [sa=eko]NUP]NP
   chicken one=CLF

   b. [[sa=eko]NUP manu]NP
      one=CLF chicken
      ‘a chicken’

Internally within the NumP, a numeral can also appear in two positions, depending on
whether the numeral is sa= ‘one’ or not. The numeral sa= ‘one’ must precede the
classifier, for example sa=eko ‘one’ (for animals), not *eko=sa. The other numerals follow
the classifier, for example eko zhua ‘two’, not *zhua eko.

A NumP can in fact contain an adjunct, we’e or su, (both) meaning ‘only’. These
adverbs however have different structural positions within the NumP. We’e must come
last, whereas su must come first within the NumP as shown in (5).

(5) a. ne ana sizha sa=ngata we’e/*su
   with child 3p one=CLF only
   ‘with their child being only one’

   b. Ito ndau dheke pe zheta wewo fata kaju ndau,
      small that climb to up up log wood that
      ngedho pe lau wena tei ne’e kazhi ko pake
      see to to down see by 3s PTCL frog
      ana no mbenu no fira, ko pake mezza
      child more lot more lot PTCL frog big.frog
      kali su/*we’e eko zhua.
      also only CLF two
      ‘The child climbed the wood and had a look behind it, and he saw
      lots of small frogs and two big ones.’

These examples have been given to illustrate the point that quantification is expressed
by numerals, and that the sortal meaning of the quantified entity is expressed by classifiers.
However, it should be noted that a classifier itself can have a quantifying function. This
kind of classifier is called a mensural classifier (cf. the classification of classifiers in Lyons
(1986: 460–466)). In other words, a sortal classifier individuates things in terms of their
kind, whereas a mensural classifier individuates things in terms of their quantity.

Rongga has a range of mensural classifiers typically associated with the quantification
of certain entities that are culturally important in daily life (e.g. corn, rice, and coconut
fruits).

Corn is counted using special terms shown in (6). A wole is the quantity of corn in one
cob. Four units of corn cobs make one mbuku. Thus, mbuku zhua ‘two mbuku’ means two
bunches of corn, with each bunch consisting of four corn cobs (i.e. 2x4). Sa=mbele is ‘one
mbele’, that is a collection of ten mbo’e, which could be 240 pieces (10x6x4).

6 The numeral sa= is the short form of the numeral esa. The classifier esa in casual speech is often
shortened to become sa. It is unclear whether the numeral esa and the classifier esa historically come from
the same source. Synchronically, they are homonyms.

7 Mbele is presumably from the Indonesian word belek ‘(big) tin can’.
Numerals in Rongga

323

(6)  
a. \( mbuku = 4 \) wole
b. \( mbo'e = 6 \) or \( 8 \) \( mbuku \)
c. \( mbele = 10 \) \( mbo'e \)

Other mensural classifiers are ‘container’ classifiers: \( neo, bondo, mbeka \) and \( tongga \). These are traditional basket-like containers woven from palm leaves of different sizes and used for carrying, transporting, or storing rice or corn.

The quantity of round objects such as eggs and coconuts is expressed by the mensural classifiers \( liwu \) and \( ulu \) (7). An example showing the use of \( liwu \) to count eggs is given in (7c). Corn can also be counted using \( ulu \) (i.e. forty pieces).

(7)  
a. \( liwu = 4 \) pieces
b. \( ulu = 10 \) \( liwu (= 40 \) pieces)
c. \( Kira-kira ko'e sa=liwu ko telo manu \)
   perhaps only one=CLF PTCL eggs chicken
   ‘There were perhaps only four chicken eggs (left).’

3.3 Form

Let us first turn to the question of formal relatedness of the numeral expressions in the two sets.

The numeral forms below ten that are related in the two sets are the expressions for ‘two’ in the restricted and regular sets respectively, \( dhua \) [\( \text{dua} \)] and \( zhua \) [\( \text{jua} \)], and for ‘one’, \( dha \) [\( \text{da} \)] and \( esa \) [\( \text{osa} \)]. These similar-looking forms appear to be Austronesian numerals, cf. PMP *duha ‘two’, *esa ‘one’ (Blust 1974; Dahl 1981, among others).

Another two familiar Austronesian forms in the two sets are \( telu \) ‘three’ and \( lima \) ‘five’ (PMP *telu ‘three’, *lima ‘five’). However, they have different distributions in each set. In the regular set these forms are used as ordinary numerals, but in the restricted set they only occur in expressions over ten, that is, \( telunggetu \) ‘thirteen’ and \( limakima \) ‘fifteen’ (1).

The other forms in the restricted set are so different from those in the regular set that they cannot have originated from the same source.

Furthermore, unlike the regular set, the restricted set does not form the numeral expressions over ten in an arithmetically consistent manner. Instead the numerals 11–17 are apparently compounds, and the first element in each of 11–15 is the corresponding 1–5 form from the regular set. The numeral \( lima-kima \) ‘fifteen’ reflects a unique rhyming pattern. The other numerals from 11 to 17 appear to contain a ligature, most often with the form \( -ngg-, \). This is unambiguous in \( esa-ngg-esa \) ‘one-LIG-one, eleven’, \( zhua-n-dua \) ‘two-LIG-two, twelve’ and \( wutu-ngg-utu \) ‘four-LIG-four, fourteen’, where each form consists of two copies of the regular numeral joined by the ligature. It is possible that \( telu-ngg-etu \) also has its origin in such a form, but that is not synchronically obvious. Certainly \( a-ngg-unae \) ‘sixteen’ and \( ne-ngg-onae \) ‘seventeen’ were not formed in this way, but may reflect synchronically unknown forms that meant ‘one-and-five’ and ‘two-and-five’.

The restricted set forms for 11–15 (with the possible exception of 13) display some kind of rhyme, giving rise to a flavour of parallelism. Parallelism is a well-known feature of

---

8 There seems to be variation in the exact quantity of a \( mbo'e \): it may contain ten \( mbukus \) (forty \( woles \) of corn).
ritual language in eastern Indonesia, and is prominent in Rongga in both ritual and non-ritual language.

The numeral expressions under and over ten in the restricted set are not formally related. This is partly because the restricted set numeral expressions for 11–15 all make use of the regular set expressions below ten.

Finally, the restricted set has no expressions for numerals higher than ‘twenty’. The regular set forms for 1–5, as noted above, reflect inherited Austronesian forms, except for wutu ‘four’. The form for ‘ten’, sambulu, is also Austronesian (PMP *puluq ‘ten’). The other forms in the regular set are discussed in §3.4 below.

Unlike the restricted set, the regular set has expressions for higher numerals over twenty, including gasu ‘hundred’ (PMP *Ratus) and riwu ‘thousand’ (PMP *ribu). Riwu, in fact, also means ‘a lot’. This could be an ‘old’ form, most likely the original meaning of riwu, because it is reported that the allegedly cognate forms in Formosan and Sulawesi languages could mean ‘an extraordinary high indefinite number’ and ‘any uncountable large quantity’ (Mills 1975:682; Dahl 1981). The list of numerals over twenty in Rongga is given in Appendix A.

3.4 System

Of the numerals in the restricted set, only those over ten display internal structure. The numeral expressions for one to ten shown in (1) are all monomorphemic and unique. We cannot therefore infer any bases for a numeral system, for example base four as in Keo (Appendix D) or base-five as in Ende (Appendix E). The forms for ten to twenty in (1) do not provide any clue to a system either, as discussed in §4.1 below. Our understanding is also constrained by the fact that expressions for numbers over twenty are absent.

Languages with limited expressions for counting may make use of body-part nouns (e.g. little finger, thumb, and upper arm) for this. Such systems are called ‘body part tally systems’, and are common in Papuan languages (Lean 1992; Laycock 1975; Vries 1993; Vries and Vries-Wiersma 1992). For example, in Fasu, one is meno ‘little finger’, five is kakorea ‘thumb’, six is nama ‘palm of hand’, and ten is kaeyako ‘upper arm’. Different languages may assign different tallies to the same body part, for example the body part name for ‘ear’ means fourteen in Fasu but fifteen in Foe (Lean 1992). However, there is no indication that the expressions in the restricted numeral set in Rongga are part of a tally system. Except for lima, which means ‘hand’ and is borrowed from the regular set, no other numeral terms in this set are body part terms.

The regular numeral set in Rongga, however, shows a clear system with base five (quinary) and base ten (decimal). The bimorphemic structure of the numeral expressions provides us with the counting unit involved in the system.

A base five system is visible in limaes a ‘six’ and limazhua ‘seven’. They involve arithmetic operations of addition, logically expressed by the function of five, one and two. Thus lima-es a ‘six’ is the sum of five and one, and limazhua ‘seven’ is the sum of five and two, where five is the base (or augend in Greenberg’s terminology (1978)).

---

9 The forms for sixteen and seventeen, discussed in §3.3, may have once had internal structure, but are synchronically monomorphemic.
Rongga also employs a base ten system, a system that is widespread across Austronesian languages. This is exemplified by sambulu ‘ten’, which is analysable as sa ‘one’ and mbulu ‘ten’. Intuitively, this is interpreted as one pile of ten things, that is, 1x10. Mbulu zhua ‘twenty’ is intuitively interpreted as two piles of ten things. They can be represented as following the logic expressed in numeral forms (1x)10 and 10(x2), respectively. The number mbulu ‘ten’ is the multiplicand and sa ‘one’ and zhua ‘two’ are multipliers (Greenberg 1978).

The formation of higher (complex) numerals employs a base ten system with multiplication and addition, as well as a base five system where six and seven are involved. Thus, sambulu saesa ‘eleven’ is logically ‘one’ (sa) x ‘ten’ (mbulu) + ‘one’ (sa). (Recall that the word esa in saesa is a generic classifier meaning ‘piece’.) So sambulu saesa literally means ‘one ten and a piece of one’. Likewise, zhuambulu esazhua ‘twenty two’ is ‘two’ (zhua) x ‘ten’ (mbulu) + ‘two’ (zhua); sambulu esa limazhua ‘seventeen’ is ‘one’ (sa) x ‘ten’ (mbulu) + (‘five’ (lima) + ‘two’ (zhua)), and so on.

It is not clear whether we can claim that Rongga has a base four system as found in Keo (Baird 2002) (see Appendix D). For example, the numeral expression of ‘eight’ (zhuambutu) is still analysable as having two morphemes zhua and mbutu (or perhaps -ngg+-wutu), which are related to the expressions meaning ‘two’ and ‘four’ — originally involving multiplication of two and four. However, the expression of zhuambutu is no longer considered as polymorphemic by native speakers of Rongga, and this is the only example that reflects a base four system.

The expression taraesa ‘nine’ certainly contains the word esa ‘one’. In contemporary Rongga, however, it is not considered bimorphemic, nor is it synchronically analysable as involving a subtraction (i.e. ‘ten minus one’). The available meanings of the word tara — ‘toe’, ‘small branch’, ‘reason’ — do not help us to derive the meaning ‘nine’ in its combination with esa. However, tara in Lio, a language east of Rongga, means ‘lack of, short of’ (Aron Mbete pers. comm.). Thus, tara-esa once meant ‘one short (of ten)’, that is ten minus one.

4 Discussion

4.1 No system in the restricted set?

Strictly speaking, the restricted numerals in Rongga cannot be said to form a ‘system’. We have already noted that the numerals 1–10 in (1) are unique numeral expressions. They involve no arithmetical operation.

There is no clear arithmetical operation that can be reasonably inferred from the numeral expressions in the restricted set. Esanggesa ‘eleven’ clearly contains the segment meaning ‘one’, and zhuandua ‘twelve’ contains the element meaning ‘two’. Both can be literally glossed as ‘one one’ and ‘two two’ respectively. While esanggesa could be logically understood as forming ‘eleven’ — for example, ‘one-one’ for ‘eleven’ — it is hard to conceive how ‘twelve’ of zhuandua could be logically derived from ‘two-two’ (‘22’) on the basis of what we know from ‘one-one’ in esanggesa ‘eleven’. Furthermore, esanggesa ‘eleven’ cannot be inferred to involve addition of ‘one’ to ‘ten’ because ‘ten’ is nguru in the restricted set, mbulu in the regular set.

Subtraction appears to be involved in watupesa ‘nineteen’. But watu is not ‘twenty’, nor is watup an attested morpheme. The word watu is attested in Rongga, meaning ‘stone’ — a word whose meaning is unrelated to the numeral. Other apparent morphemes that form
parts of restricted numerals are also attested as words (e.g. *kima* in *limakima* ‘fifteen’ and *ila* in *soroila* ‘eighteen’), but again they bear unrelated meanings: *kima* ‘snail’ and *ila* ‘bamboo’ (for flute).

### 4.2 Questions and speculations

Two questions remain: Why did the restricted set cease to exist? Does the numeral set provide us with any clue about Rongga’s affiliation within the Austronesian family in the area? The first has to do with language contact and endangerment, and the second with subgrouping. The second question will be discussed first. However, given our limited knowledge of Rongga and the history of Rongga and other Flores languages at this stage, no definite answers can be offered. The discussions below are based on fragmentary facts, which should be taken as tentative and preliminary. Definitive answers require further research.

#### 4.2.1 The Austronesian connection

There is no doubt that Flores languages belong to the Austronesian family. We have seen that the Rongga numeral words *(e)sa* ‘one’, *zhua* ‘two’, *telu* ‘three’, *lima* ‘five’, *mbulu* ‘ten’, *gasu* ‘hundred’ and *riwu* ‘thousand’ are clearly Austronesian (cf. the discussion on numerals in Dyen (1947), Dahl (1976, 1981), Blust (1974), and Li (2006), among others). One might question whether the numeral expressions could also provide us any clue about Rongga’s affiliation to other Austronesian languages in Flores.

Before we come to this point, it is necessary to briefly review previous historical linguistic research on Flores languages. Historical research specifically on Flores languages is reported in Fernandes (1996). Other works that also discuss Flores languages (though not exclusively) include Esser (1938), Dyen (1965), Blust (1993), and Donohue and Grimes (2008). Unfortunately none mentions Rongga (and its other small neighbouring languages in central Flores). While it has commonly been accepted that Flores languages belong to the Central Malayo-Polynesian (CMP) subgroup, the precise genealogical position of Flores languages in relation to other languages outside Flores is far from clear. Esser includes Flores languages together with Bimanese in Sumbawa, languages of Sumba, and Sawu in his Bima-Sumba group. Dyen (1965) assigns Sikka, a language in eastern Flores, to his Moluccan Linkage, but Blust (1993:263) (while admitting the problem of reconstructing the histories of languages that form a linkage as defined in Ross 1988:8) claims that there is evidence for an internal division of CMP languages into those of the lesser Sundas (which includes Flores) and those of the Moluccas.

Fernandes (1996) argues that Flores languages are legitimately grouped together as a subgroup within the CMP group. In his family grouping, Flores languages split into two groups, Eastern Flores Languages (Sikka, Lamaholot) and Western Flores languages, with the latter further subgrouped into the Komodo-Manggarai-Rembong subgroup and Ngadha-Lio-Palu’e subgroup.

---

12 Donohue and Grimes (re-)examine the phonological, semantic, and morphological innovations proposed for CMP and conclude that there is no convincing evidence to support the CMP subgroup.
My preliminary data (based on the 200 Swadesh list) shows that Rongga shares around 69 percent cognates with Ngadha, in contrast to 43 percent with Manggarai. This is an indication that Rongga is more affiliated with Ngadha than with Manggarai.

As for numerals, there are two points to note. Firstly, the forms of the numerals in Rongga are more similar to those of the languages east of Rongga. For example, the numeral [jua] ‘two’ has the cognate [zua] in Ngadha, and [jua] in Ende, and [sua] in Manggarai. The distance of the difference between [j] (Rongga) and [z] (Ngadha) is greater than that between [j] and [s] (in terms of manner of articulation and voicing). (There is a regular sound correspondence of [j] – [z] between Rongga and Ngadha.)

In addition, the syntax of the numeral expressions also shows the affiliation of Rongga with its neighbours to the east. Thus, ‘twenty’ in Rongga is [mbulu jua] with the multiplicand coming before the multiplier. It is [suampulu] or [suapulu] in Manggarai, where the multiplier comes before the multiplicand.

Secondly, the regular numeral set in Rongga also suggests that Rongga is more affiliated with its neighbouring languages to the east, that is Ngadha-Lio, than with its neighbours to the west (Manggarai). The regular set in Rongga employs base five (quinary) and base ten (decimal). The quinary basis is encountered in other languages east of Rongga such as Ngadha and Keo, but not in Manggarai (see Appendices B–D). Forms cognate with the expressions for wutu ‘four’ and taraesa ‘nine’ are also encountered in Ngadha, and even in languages further to the east such as Keo (Baird 2002) and Ende (see Appendix E). These words, or their equivalent forms, are part of a numeral set not found in Manggarai.

Whether Rongga indeed forms a subgroup with Ngadha and Lio, and whether Ferndandes’ claim of a Ngadha-Lio subgroup is well founded, are matters for further research.

4.2.2 Language contact and endangerment

Comrie (2005) points out that numeral sets are particularly susceptible to sociolinguistic changes in contact situations which, over time, results in their extinction.

Languages in Flores and the surrounding areas have been in constant contact with each other. Speakers of smaller languages are typically multilingual. Thus, a Rongga speaker speaks his/her own language, the neighbouring languages of Waerana, Manus, and the regional language of Manggarai, in addition to the national language of Indonesia. Speakers are often aware of words of their own language, and can generally point out that certain words do not belong to their language but to the neighbouring language(s). Contact between a minority language and a more dominant language, or languages, often results in the marginalisation and ultimate disappearance of the smaller language. The marginalisation of Rongga has been discussed in Arka (2005).

The restricted numeral set in Rongga described in this paper is almost extinct now. The restricted numerals are now only vaguely known by a few people of the older generation, now in their 60s. In the absence of any other (documented) evidence, and with the Rongga

---

13 Unlike Rongga, Keo is also reported to have a base four system, e.g. diwu ‘four’ and ha diwu ha esa ‘a four and a one = five’ (Baird 2002).

14 Arka (2005) discusses other evidence (mutual intelligibility, phonology, and grammar) to support the idea that Rongga is more affiliated with Ngadha/Lio than with Manggarai.
elders not remembering the details, we may never be able to answer the intriguing question about the irregular nature of the restricted numeral set.

It was suggested earlier in this paper that the restricted numerals were used in a game, but the game they were associated with is long gone. It is unclear why the game ceased to be played. We will perhaps also never know the nature of the game. When games are no longer practiced, the vocabulary associated with them also disappears. More intriguing, however, is the question of whether the restricted set had functions beyond games, as we would expect, and what these functions were. Here we have no evidence.

In modern Indonesia, the future of the regular numeral set is by no means any safer. It is, in fact, under threat from the dominant national language, Indonesian. Numerals below five or ten can still be understood by children, but the children do not generally have an active command of them, and especially not of the higher numerals over ten. Children and adults generally use Indonesian numerals for counting in everyday life. The list of numerals presented in this paper was ‘rediscovered’ after consulting with a number of elders because not everyone remembered all of the higher numerals. During my fieldwork in 2005, students from years four and five from the local elementary schools in Kisol\textsuperscript{15} were given a test on the regular numerals, but could not answer directly on the spot. They took the questions home and had to ask their parents.

Evidently, the future of Rongga lies in what can be culled from the past, and what is left of this past lies in the institutional memory of the elders.

Appendix

A. Numerals over ‘20’ in Rongga

<table>
<thead>
<tr>
<th>Numerals in Rongga</th>
<th>Numerals in Indonesian</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mbulu zhua saesa</td>
<td>‘21’</td>
<td>(10x2)+(1)</td>
</tr>
<tr>
<td>mbulu zhua esazhua</td>
<td>‘22’</td>
<td>(10x2)+(2)</td>
</tr>
<tr>
<td>mbulu telu</td>
<td>‘30’</td>
<td>(10x3)</td>
</tr>
<tr>
<td>mbulu wutu</td>
<td>‘40’</td>
<td>(10x4)</td>
</tr>
<tr>
<td>mbulu lima</td>
<td>‘50’</td>
<td>(10x5)</td>
</tr>
<tr>
<td>mbulu limaesa</td>
<td>‘60’</td>
<td>(10x(5+1))</td>
</tr>
<tr>
<td>mbulu limahua</td>
<td>‘70’</td>
<td>(10x(5+2))</td>
</tr>
<tr>
<td>mbulu zhuambutu</td>
<td>‘80’</td>
<td>(10x(8))</td>
</tr>
<tr>
<td>mbulu taraesa</td>
<td>‘90’</td>
<td>(10x(9))</td>
</tr>
<tr>
<td>sangasu</td>
<td>‘100’</td>
<td>(1x100)</td>
</tr>
<tr>
<td>sangasu mbulu zhua</td>
<td>‘120’</td>
<td>((1x100)+(10x2))</td>
</tr>
<tr>
<td>sangasu mbulu zhua esa lima</td>
<td>‘125’</td>
<td>((1x100)+(10x2)+(5))</td>
</tr>
<tr>
<td>sangasu mbulu lima</td>
<td>‘150’</td>
<td>((1x100)+(10x5))</td>
</tr>
<tr>
<td>sangasu mbulu lima esa lima</td>
<td>‘155’</td>
<td>((1x100)+(10x5)+(5))</td>
</tr>
<tr>
<td>sangasu mbulu zhuambutu</td>
<td>‘180’</td>
<td>((1x100)+(10x8))</td>
</tr>
<tr>
<td>sangasu mbulu taraesa</td>
<td>‘190’</td>
<td>(1x100)+(10x9))</td>
</tr>
<tr>
<td>sangasu mbulu taraesa esalima</td>
<td>‘195’</td>
<td>((1x100)+(10x9)+(5))</td>
</tr>
<tr>
<td>ngasu zhua</td>
<td>‘200’</td>
<td>(100x2)</td>
</tr>
<tr>
<td>ngasu zhua mbulu lima</td>
<td>‘250’</td>
<td>(100x2)+(10x5)</td>
</tr>
<tr>
<td>ngasu zhua mbulu lima esalima</td>
<td>‘255’</td>
<td>((100x2)+(10x5)+(5))</td>
</tr>
</tbody>
</table>

\textsuperscript{15} Kisol is one place on the Trans Flores Road where Rongga has been mostly marginalised. This is the place where many non-Rongga speakers live, and the place where the Catholic seminary school is also located.
C. Numerals in Ngadha

1. ngasu zhua mbulu limazhua ‘270’ ((100x2)+(10x(5+2)))
2. ngasu zhua mbulu zuambutu ‘280’ ((100x2)+(10x8))
3. ngasu zhua mbulu taraesa ‘290’ ((100x2)+(10x9))
4. ngasu telu ‘300’ (100x3)
5. ngasu wutu ‘400’ (100x4)
6. ngasu lima ‘500’ (100x5)
7. ngasu limazhua ‘700’ (100x(5+2))
8. sariwu ‘1000’ (1x1000)
9. riwu zhua esalima ‘2005’ ((1000x2)+(5))

B. Numerals in Manggarai

1. ca [tʃa] 21. suampulu ca [suampulu tʃa]
2. sua [sua] 22. suampulu sua [suampulu sua]
3. telu [tʃu] 23. suampulu telu [suampulu tʃu]
4. pat [pat] 24. suampulu pat [suampulu pat]
5. lima [lima] 25. suampulu lima [suampulu lima]
7. pitu [pitu] 27. suampulu pitu [suampulu pitu]
8. alo [alo] 28. suampulu alo [suampulu alo]
9. ciok [tʃiok] 29. suampulu ciok [suampulu tʃiok]
10. cempulu/cepulu [tʃɛmpulu]/[tʃɛpulu] 30. telumpulu [tʃɛlumpulu]
11. cempulu ca [tʃɛmpulu tʃa]/[tʃɛpulu tʃa] 40. patmpulu [patmpulu]
12. cempulu sua [tʃɛmpulu sua]/[tʃɛpulu sua] 50. limampulu [limampulu]
13. cempulu telu [tʃɛmpulu tʃu]/[tʃɛpulu tʃu] 60. enempulu [ənempulu]
14. cempulu pat [tʃɛmpulu pat]/[tʃɛpulu pat] 70. pitumpulu [pitumpulu]
15. cempulu lima [tʃɛmpulu lima]/[tʃɛpulu lima] 80. alompulu [alompulu]
16. cempulu enem [tʃɛmpulu ənem]/[tʃɛpulu ənem] 90. ciokmpulu [tʃiokmpulu]
17. cempulu pitu [tʃɛmpulu pitu]/[tʃɛpulu pitu] 100. ceratus [tʃɛratus]
18. cempulu alo [tʃɛmpulu alo]/[tʃɛpulu alo] 200. suaratus [sua ratus]
19. cempulu ciok [tʃɛmpulu tʃiok]/[tʃɛpulu tʃiok] 1000. cesebu [tʃɛsebu]
20. suampulu [suampulu] 2000. suar eu [sua eu]

C. Numerals in Ngadha

1. esa [asə] 21. bulu zua esa [bulu dʒua asə]
2. zua [dʒua] 22. bulu zua zua [bulu dʒua dʒua]
3. telu [taʃu] 23. bulu zua telu [bulu dʒua tʃu]
4. wutu [wutu] 24. bulu zua wutu [bulu dʒua wutu]
5. lima [lima] 25. bulu zua lima [bulu dʒua lima]
7. lima zua [lima dʒua] 27. bulu zua lima zua [bulu dʒua lima dʒua]
8. ruabutu [ruabutu] 28. bulu zua ruabutu [bulu dʒua ruabutu]
9. terasa [terasa] 29. bulu zua terasa [bulu dʒua terasa]
10. sebulu [səbulu] 30. bulu telu [bulu tʃu]
11. sebulu esa [səbulu əsa] 40. bulu wutu [bulu wutu]
12. sebulu zua [səbulu dʒua] 50. bulu lima [bulu lima]
13. sebulu telu [səbulu tʃu] 60. bulu lemasa [bulu lemasa]
D. Numerals 1–10 in both Kéo numeral systems (Baird 2002)

<table>
<thead>
<tr>
<th>Number</th>
<th>Base Four System</th>
<th>Base Five and Ten System</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ha 'esa</td>
<td>ha 'esa</td>
</tr>
<tr>
<td>2</td>
<td>'esa rua</td>
<td>'esa rua</td>
</tr>
<tr>
<td>3</td>
<td>'esa tedu</td>
<td>'esa tedu</td>
</tr>
<tr>
<td>4</td>
<td>divu</td>
<td>'esa wutu</td>
</tr>
<tr>
<td>5</td>
<td>ha diwu ha 'esa</td>
<td>'esa dima</td>
</tr>
<tr>
<td>6</td>
<td>ha diwu 'esa rua</td>
<td>'esa dima 'esa</td>
</tr>
<tr>
<td>7</td>
<td>ha diwu 'esa tedu</td>
<td>'esa dima rua</td>
</tr>
<tr>
<td>8</td>
<td>divu rua</td>
<td>'esa rua mbudu</td>
</tr>
<tr>
<td>9</td>
<td>divu rua ha 'esa</td>
<td>'esa tera-'esa</td>
</tr>
<tr>
<td>10</td>
<td>divu rua 'esa rua</td>
<td>ha mbudu</td>
</tr>
</tbody>
</table>

E. Numerals in Ende

<table>
<thead>
<tr>
<th>Number</th>
<th>Numeral</th>
<th>Numeral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sa ŋsa</td>
<td>sambulu ŋsa sa ŋsa</td>
</tr>
<tr>
<td>2</td>
<td>ŋsa jua</td>
<td>sambulu ŋsa jua</td>
</tr>
<tr>
<td>3</td>
<td>ŋsa tədu</td>
<td>sambulu ŋsa tədu</td>
</tr>
<tr>
<td>4</td>
<td>ŋsa wutu</td>
<td>sambulu ŋsa wutu</td>
</tr>
<tr>
<td>5</td>
<td>ŋsa jima</td>
<td>mbulu jua</td>
</tr>
<tr>
<td>6</td>
<td>ŋsa jima ŋsa</td>
<td>mbulu jua ŋsa ju</td>
</tr>
<tr>
<td>7</td>
<td>ŋsa jima jua</td>
<td>mbulu jua ŋsa jua</td>
</tr>
<tr>
<td>8</td>
<td>ŋsa rambutu</td>
<td>səŋasu</td>
</tr>
<tr>
<td>9</td>
<td>ŋsa təra ŋsa</td>
<td>ŋasu jua</td>
</tr>
<tr>
<td>10</td>
<td>sambulu</td>
<td>səribu</td>
</tr>
</tbody>
</table>

References


—— 2009, Ritual dance and song in language documentation: vera in Rongga and the struggle over culture and tradition in modern Manggarai-Indonesia. In M. Florey


Bound roots in Balinese and Indonesian – precategorials or verbs?

ADRIAN CLYNES

1 Introduction

In Austronesian languages like Tagalog, Malay, and Fijian, evidence for the existence of lexical or syntactic categories is often either hard to find, or is ambiguous. As a result, there has often been debate as to whether in a given such language the categories ‘verb’, ‘noun’ and so on are found.2

Traditionally, evidence from syntactic and morphological distribution is used to argue either for or against the existence of lexical or syntactic categories in a given language. Foley (1991), for example, argues that some Indonesian lexical morphemes lack category information, since they never occur unbound: they must occur with affixes, or as part of a compound, before they can be inserted into the syntax. He cites as an example the bound root 

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Nouns</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>bel-ajar</td>
<td>pel-ajar-an</td>
<td>pel-ajar</td>
</tr>
<tr>
<td>‘study’</td>
<td>‘lesson’</td>
<td>‘student’</td>
</tr>
<tr>
<td>meng-ajar</td>
<td>peng-ajar-an</td>
<td>peng-ajar-an</td>
</tr>
<tr>
<td>‘teach’</td>
<td>‘teaching’</td>
<td>‘teacher’</td>
</tr>
<tr>
<td>kurang-ajar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘(be) rude’ (‘insufficient-teach/learn’)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The fact that -ajar must undergo affixation or compounding before it can surface in a grammatical context is thus taken to be evidence for its ‘precategorial’ nature. Based on such observations, the term ‘precategorial’ is often used as a synonym for ‘bound root’ in some Malay/Indonesian linguistic circles.

1 An early version of this paper was presented at the 8th annual meeting of the Southeast Asian Linguistics Society, Kuala Lumpur, July 1998. I am grateful to that audience, and also to Wayan Arka and an anonymous reviewer for valuable comments; remaining errors and infelicities are mine.

2 See, for example, the references in Foley (1998) and Himmelmann (2005).
In this paper I focus on the morphological behaviour of bound lexical morphemes, or bound roots, in Indonesian and in Balinese. Himmelmann (2005) in his survey of typological characteristics of Austronesian languages distinguishes two types of bound roots:

(a) precategorials are those bound roots which do not occur without undergoing affixation or compounding, and which are ‘without […] clear evidence that one of the possible derivations from a given root is more basic than the other one(s)’ (cf. Verhaar 1984); and

(b) morphologically or syntactically subcategorised bound roots are those ‘which clearly belong to one particular morphological or syntactic category because of the affixations they occur with’ (Himmelmann 2005:129).

Precategorial lexical bases are said to be found in Indonesian (Kridalaksana 1986 lists around 1400 such bases) and in a number of other Malayic varieties (Adelaar 1992). Others to use the term apparently in the above sense include: Prentice (1987), also writing on Malay varieties; Gil (1995) on Riau Indonesian; and Artawa (1992) and Arka (2003) on Balinese, which is relatively closely related to Malay (Adelaar 2005). Sneddon (1996) is perhaps alone in asserting that bound roots in Indonesian are verbs, though he offers no evidence for that view. The second type, morphologically or syntactically subcategorised bound roots, is said to occur in Nias and Biak. In these languages they are ‘usually […] verbal roots which obligatorily have to be marked with a person-marking prefix before they can function as predicates’ (Himmelmann 2005:129).

In §2 below I present evidence from Balinese, and in §3 from Standard Indonesian that, counter to most previous suggestions, bound lexical bases in those languages are not precategorial. Instead, affixation patterns indicate that in those languages, as in Nias and Biak, such bases are categorised as verbs: they thus belong to Himmelmann’s second type. I begin with Balinese, since its patterns are simple and clear, and perhaps help clarify what is going on in Indonesian.

2 Balinese bound lexical morphemes

In Balinese bound lexical roots surface after affixation as both nouns and verbs. To simplify the exposition, I will focus on verb affixation. Verbs generally have the structure (prefix)-root-(suffix) (cf. Clynes 1995). Some verbs surface without further affixation as free morphemes; others only occur with a prefix and/or a suffix. Formally, there are four main prefixing possibilities, as shown in Table 1.

---

3 At the level of syntax, both of these languages show reasonably clear evidence for a distinction between nouns and verbs: NP predicates cannot be negated by the verbal negator, tidak in Indonesian, sing in Balinese.

4 Anticipating the analysis presented below, nouns formed from bound roots can be safely assumed to derive from verbs.

5 To simplify the exposition, I have ignored reduplication. The behaviour of reduplicated root morphemes, free or bound, parallels that of non-reduplicated morphemes with respect to the issues discussed here. For more detail on Balinese verbal affixation, including the (morphological) split-S patterning, see Clynes (1995) and Arka (2003).
Of the four main verbal prefixes in Balinese, only two are of relevance to this study: \(N\)-, and its allomorph \(me^-1\), since they commonly attach to bound roots. The others are secondary in that they derive lexemes from other independently existing lexemes, such as in the above examples, and are not discussed in detail below.

The prefix, or lack of it, gives information about the semantic role of the subject NP. If the subject is an Undergoer, then there is no prefix; if the subject NP is an Actor, the verb takes the nasal prefix ‘\(N^\cdot\)? This is seen in the basic voice alternation on transitive verbs:

(2) a. Undergoer Voice (UV) form Actor Voice (AV) form

\[
\begin{align*}
  adep & \quad ngadep & \quad (N\text{-}adep) & \quad \text{‘sell’} \\
  jemak & \quad nyemak & \quad (N\text{-}jemak) & \quad \text{‘take’} \\
  beli & \quad meli & \quad (N\text{-}beli) & \quad \text{‘buy’} \\
  tebek & \quad nebek & \quad (N\text{-}tebek) & \quad \text{‘stab’}
\end{align*}
\]

b. \(i\) ketut nyemak pipis

\(\text{DET Ketut AV:take money ‘Ketut took (the) money.’}\)

c. pipis-\(\acute{e}\) jemak=\(a\) tekén \(i\) ketut

\(\text{money-DEF UV:took=3.ACT with DET Ketut ‘The money was taken by Ketut.’}\)

The same basic possibilities are found on intransitives:

(3) a. Undergoer subjects (prefixless):

\[
\begin{align*}
  labuh & \quad \text{‘fall (animate faller)’} & \quad ulung & \quad \text{‘fall’} \\
  pédél & \quad \text{‘squashed flat’} & \quad takut & \quad \text{‘afraid’} \\
  mati & \quad \text{‘die/dead’} & \quad selem & \quad \text{‘black’}
\end{align*}
\]

6 I follow the standard Balinese orthography, with these exceptions: /a/ is everywhere represented as \(<a>\), /e/ as \(<e>\). In the standard orthography /a/ is \(<a>\) in prefixes, \(<e>\) elsewhere; /e/ is also represented by \(<e>\).

7 \(N\)- surfaces as /\(\acute{n}\)/ before root-initial sonorants, and as a homorganic nasal ‘replacing’ root-initial obstruents. As indicated above, a subset of bound intransitives with Actor subjects takes the prefix \(me^-1\), an allomorph of /\(N\)/, e.g., \(me\text{-}laib ‘run’ me\text{-}daar ‘eat’, me\text{-}siram ‘bathe’. See Clynes (1995) and Arka (2003) for more detailed analyses.
b. Actor subjects (with N-):

- Negak (N-togak) ‘sit’
- Meju (N-paju) ‘defecate’
- Ngigel ‘dance’
- Nylémpoh (N-jlemphoh) ‘fall to one’s knees’
- Ngelur (N-golur) ‘yell’
- Ng-wél ‘complain angrily’

To summarise relevant verb prefixation in Balinese, if the subject NP is an Actor, it is signalled on the verb by a prefix, and if it is an Undergoer, it is signalled on the verb by the absence of a prefix.

Suffixes occur on both intransitive and transitive verbs. Only two suffixes are found on intransitive verbs, both of the form -an. Both generally function to derive intransitives from bases of another syntactic category: (1) the suffix -anj derives intransitive verbs from nouns, with the meaning ‘be covered with/in [noun]’, e.g. semut ‘ant’ > semut-an ‘be covered in ants’; (2) -an2, (an allomorph of both -in and -ang) derives static intransitive verbs with Undergoer subjects from transitives, with meaning ‘be in a state of having been [verbed]’, and almost always occurs in combination with me-2, e.g. m-ubet-an ‘be open (e.g. a door)’. A third suffix, -anj3, found on nouns, attaches to verbal roots. The derived noun has the meaning ‘thing which undergoes the action of [verb]’, e.g. tulis ‘write, V.TR’ tulis-an ‘writing, thing written’. Both -anj2 and -anj3 (but not, it seems, -anj1) attach to bound roots; for examples see Appendix 1.

On transitive verbs the suffix gives further information about the semantic role of the core Undergoer NP argument. Suffixes obligatorily occur on many transitive verbs. There is a three-way paradigmatic contrast in line with that reconstructed by Pawley and Reid (1979) for Proto Austronesian:

(4) ‘-Ø’ Undergoer is a ‘true patient’: an entity totally affected by the action of the verb, e.g. severely damaged, destroyed, or created
- in Undergoer is a ‘location’
- ang Undergoer is a ‘theme, beneficiary, causee, or other circumstantial role’

Note that just as Undergoers are the literally unmarked subject choice for intransitives, so when the Undergoers of transitive verbs are at their most prototypical, as ‘true patients’, this is signalled by a complete lack of suffixation. ‘Partially affected’ Undergoers on the other hand take either -in or -ang. Examples of all three types of transitive verbs in Balinese are given in (5).

(5)

<table>
<thead>
<tr>
<th>‘UG = true patient’ -Ø</th>
<th>‘UG = location’ -in</th>
<th>‘UG = theme, beneficiary, etc.’ -ang</th>
</tr>
</thead>
<tbody>
<tr>
<td>jagur ‘punch’</td>
<td>oot-in ‘unload from UG’</td>
<td>anti-ang ‘wait for UG’</td>
</tr>
<tr>
<td>godot ‘cut up’</td>
<td>usu-usu-in ‘lightly stroke UG’</td>
<td>beli-ang ‘buy for UG’</td>
</tr>
<tr>
<td>beli ‘buy’</td>
<td>ubet-in ‘open door to UG’</td>
<td>ubet-ang ‘open UG; open [theme for UG]’</td>
</tr>
<tr>
<td>gaé ‘make’</td>
<td>ulap-in ‘wave to UG’</td>
<td>tegak-ang ‘cause UG to sit’</td>
</tr>
</tbody>
</table>

---

8 Another subset of the intransitives, with Undergoer subjects, take a distinct prefix me-2; m-adep ‘be sold’, me-jemak ‘be taken’. These are clearly derived from transitive bases, cf. Clynes (1995:242–246).
Relevant verbal affixation patterns in Balinese are summarised in (6). As in many Austronesian languages, verbs obligatorily carry affixes to signal information about the semantic roles of core nominal arguments:

(6) (i) All verbs must carry a prefix (N-) if they have Actor subjects.
    (ii) Verbs with Undergoer subjects carry no prefix (‘Ø-’).
    (iii) All transitive verbs must carry a suffix if they have only ‘partially affected’ Undergoers (location, theme, beneficiary, causee, or other circumstantial role).

It follows that there are only two possible cases in Balinese where verbs do not carry affixing, namely: (1) Intransitive verbs with Undergoer subjects, like pules ‘sleep’, gadang ‘(be) green’, and (2) Transitive verbs in Undergoer Voice, where the Undergoer is a true patient, like godot ‘slice’, cakcak ‘chop up’.

Many other verbs are, therefore, like oot-in ‘unload from UG’, usu-usu-in ‘lightly stroke UG’, ulap-in ‘wave to UG’, ubet-ang ‘open (door/window)’ and anti-ang ‘wait for UG’ in (5), or the agentive intransitives listed in (3b) in that the bases from which they are formed never occur without affixation. Moreover, many such bases have no readily elicitable nominal form. They are apparently inherently verbal, yet must surface bearing suffixes due to their inherent semantic-syntactic makeup.

This then is our hypothesis as to why bound lexical morphemes are found in Balinese — not because they lack category status, but because they are verbs, verbs whose inherent semantic-syntactic content is such that it is obligatorily signalled by affixation. This analysis makes three unambiguous, testable predictions about bound lexical morphemes in Balinese:

(7.1) All bound lexical morphemes will have at least one lexical realisation taking a ‘verbal’ affix. (So, as a corollary, no bound lexical morpheme will occur only with affixes deriving categories other than verbs, such as affixes deriving nouns, to the exclusion of verbs).

(7.2) That lexical realisation will be an ‘agentive’ verb, either: an agentive intransitive verb (taking N-); or a transitive verb, and so subcategorised for an Actor NP (as well as an Undergoer).

(7.3) In their semantic behaviour, when they combine with affixes bound lexical morphemes will pattern like free (transitive or intransitive) verbs, and not like free morphemes of other classes such as nouns.

If, on the other hand, these morphemes are precategorial, having no orientation with respect to syntactic class, we get a very different set of predictions:

(8.1) Bound lexical morphemes will show random affixation possibilities — probably the majority will occur with both nominal and verbal affixes, though it may be that some will occur with nominal affixes only, and some with verbal affixes only.

(8.2) In their semantico-combinatorial behaviour, bound lexical morphemes will pattern randomly, not like free verbs, and not like free nouns.
Even an impressionistic survey of Kamus Bali-Indonesia (Panitia Penyusunan Kamus Bali 1990), the principal dictionary of Balinese, indicates that predictions (7.1) to (7.3) hold very strongly. Table 2 summarises patterns found in a sample of 100 bound lexical morphemes taken from that source, and presented in Appendix 1. Agreement or otherwise with predictions (7.1), (7.2), (7.3) and (8.1) is indicated.⁹

**Table 2: Affixation patterns of bound lexical morphemes in Balinese**

<table>
<thead>
<tr>
<th>Root type</th>
<th>Affixes</th>
<th>Total</th>
<th>Percentage</th>
<th>Consistent (✓) or inconsistent (✗) with prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Roots taking nominal affixes only</td>
<td>-an³, RDP-an, pe-….an</td>
<td>0</td>
<td>0%</td>
<td>(7.1)✓, (7.2)✓; (8.1)✗</td>
</tr>
<tr>
<td>(b) Roots taking verbal affixes only</td>
<td>N-, me-1, -ang, -in, me-…-an², me-RDP-an², -an² UG, pe-…-in, pe-…-ang, RDP</td>
<td>64</td>
<td>64%</td>
<td>(7.1)✓; (8.1)✗</td>
</tr>
<tr>
<td>(c) Roots taking verbal and nominal affixes</td>
<td></td>
<td>36</td>
<td>36%</td>
<td>(7.1)✓; (8.1)✓</td>
</tr>
<tr>
<td>(d) Roots where at least one of the surfacing verb forms has an agentive subject</td>
<td></td>
<td>93</td>
<td>93%</td>
<td>(7.2)✓</td>
</tr>
</tbody>
</table>

That 100% of the sample occurs with at least one verbal affix is consistent with prediction (7.1). That no roots occurred with a nominal affix only is as predicted by (7.2). The underlyingly verbal nature of the bound lexical morphemes is strongly supported. The patterning is clearly not random, counter to prediction (8.1), and so not that which would be expected if the root morphemes were precategorial. That 93% of the sample had a ‘derived’ verb form with an agentive subject is consistent with prediction (7.2).

Prediction (7.3) requires that bound lexical morphemes will pattern like free verbs (either transitive or intransitive) and not like free nouns in their semantic and syntactic behaviour when bearing ‘derivational’ affixes. This is confirmed by patterns of verbal affixation with the applicative affix -ang, the nominal-deriving affix -an, and the locative affix -in. I present the evidence in data sets (9), (10) and (11) respectively.

When occurring with -ang, free intransitive verb bases derive verbs with a causative sense (9a):

(9) a. [Free V.INTR]-ang > ‘ACT causes UG to [V.INTR]’ (UG = S of V.INTR)
    *pules ‘sleep’*                  *pules-ang*  ‘cause UG to sleep’
    *ulung ‘fall’*                   *ulung-ang*  ‘cause UG to fall’

---

⁹ The 100-item sample was made up of the first 20 bound lexical morphemes listed in the sections for each of the letters a, b, c, d and e. See Appendix 1.
As exemplified by data set (3), many bound lexical morphemes occur as agentive intransitives when affixed with N- or its allomorph me-. Such bound roots behave precisely like free intransitives when they occur with -ang, in also forming verbs with a causative sense (9b):

(9) b. [Bound V.INTR]-ang\textsuperscript{10} > ‘ACT causes UG to [V.INTR]’ (UG = S of V.INTR)  

tegak ‘sit’  
siram ‘bathe:HON’  

-ang  
tegak-ang ‘cause UG to sit’  
siram-ang ‘cause UG to bathe’

When free transitive verb bases occur with -ang they derive verbs with a benefactee, recipient or other applicative-type role as the Undergoer argument (9c):

(9) c. [Free V.TR]-ang > ‘ACT does [V.TR] for UG’  

jagur ‘punch’  
beli ‘buy’  

-ang  
jagur-ang ‘punch [theme] for UG’  
beli-ang ‘buy [theme] for UG’

Certain bound roots exhibit identical behaviour, and are accordingly assumed here to be underlyingly transitive verbs (9b):

(9) d. [Bound V.TR]-ang > ‘ACT does [V.TR] for UG’  

ubet- ‘open’  
anti- ‘wait’  

-ang  
ubet-ang ‘open [theme] for UG’  
anti-ang ‘wait for UG’

On the other hand, free morphemes occurring in NP argument slots (nouns) do not generally occur with -ang (9e). They thus exhibit behaviour quite different from that of bound lexical roots (see also Table 2, part (a)):\textsuperscript{11}

(9) e. [Noun]-ang

sepatu ‘shoe’  
papur ‘powder’  

-ang  
*sepatu-ang (does not occur)  
*papur-ang (does not occur)

The same remarks apply in relation to the nominal-deriving affix -an: with bound bases (10a), the derivational pattern parallels that of free verbs (10b), rather than nouns (10c). Both bound bases and free verbs combine with -an to form nouns with a locative meaning ‘thing/place on/at/to/towards which the action of [verb base] is done’, whereas most free nouns simply do not occur with -an. The small number which do form stative verbs with a very different meaning, ‘be covered with/in [noun]’ (10d):

(10) a. -tegak ‘sit, V.INTR’  

tegak-ang ‘thing sat on’

-siram ‘bathe:HON, V.INTR’  
siram-ang ‘place bathed at’

b. pules ‘sleep’  

demen ‘like, V.INTR’

pules-ang ‘thing slept on’  
demen-ang ‘thing which is liked’

c. sepatu ‘shoe’  
papur ‘powder’

*sepatu-ang (does not occur)  
*papur-ang (does not occur)

\textsuperscript{10} Consistent with the behaviour evidenced here, these bound roots are glossed as underlyingly (agentive) intransitive verbs, cf. negak (N-tegak) ‘sit, V.INTR’, and me-siram ‘bathe:HON, V.INTR’

\textsuperscript{11} A small subset of free nouns does take -ang to produce verbs meaning ‘use [root] as term of address’ — not the meaning found when bound lexical morphemes take -ang.
The predicted pattern again occurs with the locative verbal affix -in (11a)–(11c). Both free verbal roots and bound lexical morphemes form verbs with the general meaning “ACT does [V.INTR/V.TR/action of bound base] on/at UG”. When the base is a free noun, the derived meaning is very different, that is, “ACT applies [noun] to UG” (11d):

(11) a. [Free V.INTR]-in  ‘ACT does [V.INTR] on/at UG’
   pules  ‘sleep’  pules-in  ‘sleep on UG’
   ulung  ‘fall’  ulung-in  ‘fall on UG’

b. [Free V.TR]-in  ‘ACT does [V.TR] on/at UG’
   jagur  ‘punch’  jagur-in  ‘hit at [theme] repeatedly’
   beli  ‘buy’  beli-in  ‘buy [theme] at/from UG’

c. [Bound V.INTR]-in  ‘ACT does [V.INTR] on/at UG’
   negak (N-tegak)  ‘sit’  tegak-in  ‘sit on UG’ (not *‘put a seat on UG’,
   me-siram  ‘bathe.HON’  siram-in  ‘bathe in UG’ (not *‘put a bath on UG’
   cf. (11d))

d. [Free Noun]-in (LOC)  ‘ACT applies [Noun] to UG’
   sepatu  ‘shoe’  sepatu-in  ‘put shoe on UG’
   pupur  ‘powder’  pupur-in  ‘put powder on UG’

The regular parallels in semantic-combinatorial behaviour demonstrated in (9), (10) and (11) are exactly those predicted by (7.3). They are consistent with the hypothesis that bound lexical morphemes in Balinese are verbs, and not nouns or precategorials. The patterns in (9) to (11) are not those predicted by the precategoriality hypothesis (cf. (8.1), (8.2)). Moreover, from patterns like those in (9a) and (9b) it is possible to conclude that -tegak and -siram (and similar bases such as those in data set (3b)), are underlyingly intransitive verbs and not simply verbal. Similarly the evidence in (9d) indicates that other bound roots (such as those taking -in in column 2 of data set 5) pattern like transitive verbs. These bound lexical bases thus carry information about morphosyntactic category and argument structure.

Both the combinatorial patterns of bound lexical morphemes (Table 2), and the semantic-syntactic behaviour of the resulting affixed forms confirm the hypothesis that bound lexical morphemes in Balinese are verbs (Clynes 1995). The combinatorial facts follow from the requirement in Balinese that semantic information about the agency or ‘patiency’ of core nominal arguments be marked on the verb. In Balinese, semantically agentive underlying intransitives will therefore not be able to surface without N-, while underlying transitives with only ‘partially affected’ Undergoer arguments will not be able to surface without either -in or -ang.

12 Where underlying transitive ‘verbs of violence’ take -in, the UG is a location/surface and the action is repeatedly directed at that surface. The UG may thus be less affected than the U of the same verbal base when it occurs without a suffix, see Clynes (1995).
Given that affixing patterns similar to those in Balinese are found in many western Austronesian languages\(^{13}\) it is reasonable to hypothesise that in at least some related languages bound lexical morphemes will also pattern as verbs. Further, given a similarly widespread tendency for ‘Patient Primacy’, or the assigning of unmarked status to the Undergoer (rather than Actor) arguments, one can predict that often those bound verbs will surface with affixes signalling the presence of Actor arguments. Predictions (7.1) to (7.3) should therefore hold for a number of other Austronesian languages, where the simplest analysis will similarly be that bound roots are verbs.

In the rest of this paper predictions (7.1) and (7.2) are tested on data from Indonesian, a variety of Malay. Those predictions are strongly confirmed.

3 **Bound lexical morphemes in Indonesian**

The analysis argued for Balinese above is independently proposed for Indonesian by Sneddon: ‘[a] base which cannot occur alone as a word is a verbal base’ (1996:60). Sneddon does not, however, produce evidence in support of his claim. To test the hypothesis, I examined a sample of the first 200 items in the list of approximately 1400 ‘precategorials’ cited by Kridalaksana (1986). For each of the bound roots, the meaning and affixation possibilities were noted, as recorded in the major dictionary of Indonesian, the *Kamus Besar Bahasa Indonesia* (Pasat Pembinaan dan Pengembangan Bahasa 1990). The detailed counts are given in Table 3 and the full data in Appendix 2.

### Table 3: Affixation patterns of bound lexical morphemes in Indonesian (1)

<table>
<thead>
<tr>
<th>Root type</th>
<th>Affixes attaching to root</th>
<th>Total</th>
<th>% of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Roots taking only nominal affixes</td>
<td>-an/per/-per-...-an/peng/-peng-...-an/RDP -an, etc.</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Sub-total</td>
<td></td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>(b) Roots taking only verbal affixes</td>
<td><em>meng-</em></td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>meng-</em> , ter-*</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ber-</em> (AGT)(^{14})</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ber-</em> (AGT), ter-*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>meng-</em> , ber-* (AGT)</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>meng-</em> , ber-* (AGT), ter-*</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ter-</em></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>ber-</em> (non-AGT)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Sub-total</td>
<td></td>
<td>145</td>
<td>72.5%</td>
</tr>
</tbody>
</table>

\(^{13}\) Compare Pawley and Reid (1979) and the three-way suffixing contrast in (4) above.

\(^{14}\) Note: AGT = subject is an Actor/Agent.
The data in Table 3 conform closely to predictions (7.1) and (7.2). All 200 items in the sample have at least one lexical realisation taking a verbal affix (prediction (7.1)). Prediction (7.2) is confirmed to the extent that 175 of the roots surface with the agentive prefix meng-, whose affixational possibilities and semantics closely parallel those of Balinese N-,

while another 10 verbs occur only with the ter- prefix, which attaches to verbs normally or potentially having a controlling, animate actor. Ter- functions to override that reading, marking accidental performance of, or inability to perform, the action. Of the 15 remaining roots in the sample, 12 take the less-well understood ber-prefix, but have unambiguously agentive subjects (cf. Appendix 2). A further three verbs taking ber- have non-agentive subjects. If we count simply verbs with agentive subjects (including the 12 with ber-), they make up 197 of 200, or 98.5% of the sample. These patterns are summed up in Table 4.

Table 4: Affixation patterns of bound lexical morphemes in Indonesian (2)

<table>
<thead>
<tr>
<th>(a) Roots taking meng-</th>
<th>meng- only</th>
<th>85</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>meng-, ter-only</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>meng- + -an (NOM) +/- other</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>meng-, ber- + nominal affixes</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>meng-, ber- (non-AGT) + nominal affixes</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>meng-, ber- (AGT)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>meng-, ber- (AGT), ter-</td>
<td>5</td>
</tr>
<tr>
<td>Sub-total</td>
<td>175</td>
<td>87.5%</td>
</tr>
<tr>
<td>(b) Roots taking ter-</td>
<td>ter- only</td>
<td>10</td>
</tr>
<tr>
<td>(c) Roots taking ber- only, with agentive subject</td>
<td>ber- AGT only</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>ber- AGT, ter-only</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Running total, agentive subjects</td>
<td>197</td>
</tr>
<tr>
<td>(d) Roots taking ber-only, with non-agentive subject</td>
<td>ber- non-AGT only</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>200</td>
<td>100%</td>
</tr>
</tbody>
</table>

Refer to data sets (2) and (3) and discussion there.
The statistics in Tables 3 and 4 thus closely conform to the predictions in (7), which require bound roots to be underlyingly verbs, and to generally occur with agentive subjects.

In contrast, predictions (8.1) and (8.2) are predicated on underlying precategorial status, and should be true if bound verbs are underlyingly precategorial. Prediction (8.2) was not tested for Indonesian (though it was for Balinese and was found to fail for that language, see discussion of (9)–(11)), but prediction (8.1) clearly fails. Affixation patterns are not random, but are instead very strongly skewed towards ‘verbness’: 100% of the sample takes verbal affixes, including 72.5% which takes only verbal morphology. No bound lexical morphemes (0%) take only nominal affixes, and only 27.5% take any nominal affixation at all.

We can conclude from the patterns in Tables 3 and 4 that in Indonesian, bound lexical morphemes obligatorily surface as verbs, not as other categories. They are therefore most plausibly analysed as being listed in the lexicon as verbs. The ‘alternative’, that bound lexical morphemes are precategorial but are listed as ‘obligatorily surfacing as verbs’, is not in fact an alternative position, but simply a restatement of the same analysis: bound lexical morphemes are lexically listed as verbs in Indonesian.

4 Conclusions

The main hypothesis of this study, that bound lexical morphemes in Balinese and Indonesian are verbs, is supported by all the evidence considered above. This finding in turn throws doubt on the precategoriality hypothesis often put for the Malayic languages, and for relatively closely related languages like Balinese. It would be interesting to extend the analysis to related Austronesian languages such as Javanese, to see whether the strong predictions in (7) hold for those languages. A reasonable expectation based on the present findings is that they will.

The data above perhaps suggest the need for a more nuanced understanding of the function of affixation in such languages. Verbal affixation is often said to be purely derivational (or almost so) in Malay/Indonesian, Balinese and other western Indonesian languages. Yet in the case of bound intransitive verbs at least, the affixes they carry appear to function in a similar way, say, to subject agreement inflection in English, signalling inherent information about, or constraints on, syntactic and semantic features of core NPs. In such cases verbal affixation seems close to having an iconic or inflectional function, rather than deriving one thing from another.

Appendix 1

Sample of Balinese bound roots, from Panitia Penyusunan Kamus Bali (1990). First 20 items from each of the sections a, b, c, d, and e. The meaning given in the rightmost column is that of the first-listed form in the ‘Verbal affixes’ column. ‘AGT’ indicates the subject of the verb is agentive and ‘UG’ indicates the subject of the verb is Undergoer-like.
<table>
<thead>
<tr>
<th>Root</th>
<th>Verbal affixes</th>
<th>Nominal affixes</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> aat</td>
<td>N- V.INTR AGT; -ang AGT</td>
<td>RDP-an ‘thing desired’</td>
<td>‘desire’</td>
</tr>
<tr>
<td><strong>2</strong> abab</td>
<td>-in AGT; me-…-an</td>
<td>–</td>
<td>‘stir to remove dregs’</td>
</tr>
<tr>
<td><strong>3</strong> abang</td>
<td>-in AGT</td>
<td>-an ‘drainage channel’</td>
<td>‘put in a drainage channel’</td>
</tr>
<tr>
<td><strong>4</strong> aben</td>
<td>N- ‘cremate’ V.INTR AGT; me-UG; -ang AGT</td>
<td>peng-…-an ‘cremation’</td>
<td>‘cremate’</td>
</tr>
<tr>
<td><strong>5</strong> abil</td>
<td>N-…-in AGT</td>
<td>-an ‘lump of wood’</td>
<td>‘cut down a lump of wood’</td>
</tr>
<tr>
<td><strong>6</strong> adak</td>
<td>-ang AGT</td>
<td>RDP-an ‘thing created’</td>
<td>‘create’</td>
</tr>
<tr>
<td><strong>7</strong> adon</td>
<td>N-…-ang; me-…-an</td>
<td>RDP-an ‘product of mixing with hands’</td>
<td>‘mix with the hands’</td>
</tr>
<tr>
<td><strong>8</strong> adug</td>
<td>N- V.INTR AGT; N-RDP</td>
<td>–</td>
<td>‘abuse one’s power’</td>
</tr>
<tr>
<td><strong>9</strong> adur</td>
<td>me-…-an UG; me-RDP an U</td>
<td>–</td>
<td>‘be in a mess’</td>
</tr>
<tr>
<td><strong>10</strong> agen</td>
<td>N- V.INTR AGT with clausal complement; -ang ‘hope for’</td>
<td>RDP-an ‘thing hoped for’</td>
<td>‘hope, intend’</td>
</tr>
<tr>
<td><strong>11</strong> agu</td>
<td>N- V.INTR AGT; -ang ‘act arrogantly’</td>
<td>–</td>
<td>‘arrogant’</td>
</tr>
<tr>
<td><strong>12</strong> ajag</td>
<td>N- V.INTR AGT; -in</td>
<td>–</td>
<td>‘commute’</td>
</tr>
<tr>
<td><strong>13</strong> alad</td>
<td>N- V.INTR UG? ‘long’ of sound; -ang AGT</td>
<td>–</td>
<td>‘long (of sound)’</td>
</tr>
<tr>
<td><strong>14</strong> alang</td>
<td>-in AGT; ka-pi- ‘be obstructed’</td>
<td>peng-; -an ‘thing which obstructs’</td>
<td>‘obstruct’</td>
</tr>
<tr>
<td><strong>15</strong> aling</td>
<td>me-…-an V.INTR, AGT</td>
<td>RDP; -an ‘obstructing wall’</td>
<td>‘hide’</td>
</tr>
<tr>
<td><strong>16</strong> ambiar</td>
<td>N- V.INTR, AGT; -in ‘approach, etc.’; p-…-an ADV, ‘(do) in full view’</td>
<td>–</td>
<td>‘do something (e.g. sit) in a line, while visible’</td>
</tr>
<tr>
<td><strong>17</strong> ambul</td>
<td>N- V.INTR AGT; -ang; -in; me-…-an AGT, RCP</td>
<td>-an ‘object of sulking’</td>
<td>‘sulk’</td>
</tr>
<tr>
<td><strong>18</strong> amik</td>
<td>N- V.INTR AGT</td>
<td>RDP-an ‘thing snacked on’</td>
<td>‘snack’</td>
</tr>
<tr>
<td><strong>19</strong> amput</td>
<td>-ang; -in AGT</td>
<td>–</td>
<td>‘fan out’ (-ang V.TR); ‘clean’ (-in V.TR)</td>
</tr>
<tr>
<td><strong>20</strong> anan</td>
<td>-in AGT, me-…-an</td>
<td>–</td>
<td>‘clear (edge of ricefield)’</td>
</tr>
<tr>
<td><strong>21</strong> babak</td>
<td>N-…-in AGT</td>
<td>-an ‘land which has been ploughed’</td>
<td>‘plough land to make a ricefield’</td>
</tr>
<tr>
<td><strong>22</strong> bading</td>
<td>-ang; -in AGT; me- UG; RDP</td>
<td>–</td>
<td>‘turn over’</td>
</tr>
<tr>
<td><strong>23</strong> bakal</td>
<td>N- V.INTR AGT</td>
<td>-an ‘land which has started to be ploughed’</td>
<td>‘begin to plough’</td>
</tr>
<tr>
<td><strong>24</strong> balih</td>
<td>me- V.INTR AGT; pe-; pe-…-ang</td>
<td>RDP-an</td>
<td>‘watch’</td>
</tr>
<tr>
<td><strong>25</strong> banjah</td>
<td>-ang; -in AGT; me- UG</td>
<td>–</td>
<td>‘divide, share’</td>
</tr>
<tr>
<td>No.</td>
<td>Word</td>
<td>Short Meaning</td>
<td>Long Meaning</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>26</td>
<td>baseh</td>
<td>-ang; -in AGT; me- AGT</td>
<td>pe- ~ -an ‘place to wash’</td>
</tr>
<tr>
<td>27</td>
<td>batar</td>
<td>-ang, -in AGT</td>
<td>- an (product of verb = a floor)</td>
</tr>
<tr>
<td>28</td>
<td>bedag</td>
<td>N- V.INTR AGT</td>
<td>RDP-an ‘thing hunted’</td>
</tr>
<tr>
<td>29</td>
<td>begbeg</td>
<td>N- V.INTR AGT</td>
<td>peN-an ‘tool/trap for’</td>
</tr>
<tr>
<td>30</td>
<td>bejadi</td>
<td>-in AGT; -an UG</td>
<td>–</td>
</tr>
<tr>
<td>31</td>
<td>beki</td>
<td>-nin AGT; be- V.INTR, ‘enjoy annoying’</td>
<td>–</td>
</tr>
<tr>
<td>32</td>
<td>belbel</td>
<td>-an UG; N-..-in CAUS, AGT; ke-..-an</td>
<td>–</td>
</tr>
<tr>
<td>33</td>
<td>belet</td>
<td>ke- UG, -ang, -in AGT; me-RDP-an AGT, RCP</td>
<td>–</td>
</tr>
<tr>
<td>34</td>
<td>benah</td>
<td>-ang, -in AGT</td>
<td>–</td>
</tr>
<tr>
<td>35</td>
<td>bered</td>
<td>me- V.INTR AGT; -ang ‘form into a line’</td>
<td>–</td>
</tr>
<tr>
<td>36</td>
<td>besah</td>
<td>-ang, -in AGT</td>
<td>-an ‘thing washed’</td>
</tr>
<tr>
<td>37</td>
<td>biayag</td>
<td>-ang, -in AGT, me-..-an UG</td>
<td>–</td>
</tr>
<tr>
<td>38</td>
<td>binder</td>
<td>me- V.INTR AGT; -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>39</td>
<td>blesat</td>
<td>me- V.INTR AGT; pe- V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>40</td>
<td>belet</td>
<td>-ang, -in AGT; ke-; me-..-an</td>
<td>–</td>
</tr>
<tr>
<td>41</td>
<td>cabcab</td>
<td>-in AGT</td>
<td>-an ‘waste water, used water’</td>
</tr>
<tr>
<td>42</td>
<td>canda</td>
<td>me- V.INTR AGT; -in AGT, ke-in</td>
<td>–</td>
</tr>
<tr>
<td>43</td>
<td>candet</td>
<td>N- V.INTR AGT; me-..-an</td>
<td>–</td>
</tr>
<tr>
<td>44</td>
<td>cangkrim</td>
<td>me-..-an V.INTR AGT</td>
<td>RDP-an ‘riddle sung as a lullaby’</td>
</tr>
<tr>
<td>45</td>
<td>capat</td>
<td>-in AGT; me-..-an V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>46</td>
<td>cavis</td>
<td>me- UG; -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>47</td>
<td>cebélég</td>
<td>N- UG?; me-; pe-..-RDP</td>
<td>–</td>
</tr>
<tr>
<td>48</td>
<td>cégélèh</td>
<td>me- V.INTR AGT, N- V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>49</td>
<td>cékcék</td>
<td>N- AGT; me-..-an AGT</td>
<td>–</td>
</tr>
<tr>
<td>50</td>
<td>celed</td>
<td>-in AGT; me-..-an</td>
<td>-an ‘threads which have been tied’</td>
</tr>
<tr>
<td>51</td>
<td>celek</td>
<td>-ang AGT</td>
<td>pi-, pe- ‘emphatic word’</td>
</tr>
<tr>
<td>52</td>
<td>celep</td>
<td>me- V.INTR AGT; -in, -ang AGT; RDP</td>
<td>–</td>
</tr>
<tr>
<td>53</td>
<td>cempléng</td>
<td>me- AGT, RDP</td>
<td>–</td>
</tr>
<tr>
<td>54</td>
<td>cénél</td>
<td>N- V.INTR AGT; -ang</td>
<td>-an ‘place where one hooks s.th. onto s.th.; snag’</td>
</tr>
<tr>
<td>55</td>
<td>cepol</td>
<td>me- UG?: -ang, -in AGT, RDP, pe- ~ -ang</td>
<td>–</td>
</tr>
<tr>
<td>56</td>
<td>cingcing</td>
<td>-ang AGT; me-...-an</td>
<td>–</td>
</tr>
<tr>
<td>57</td>
<td>claduk</td>
<td>pe- V.INTR, me- , -ang AGT, me-...-an</td>
<td>–</td>
</tr>
<tr>
<td>58</td>
<td>clegos</td>
<td>me- UG, RDP</td>
<td>–</td>
</tr>
<tr>
<td>59</td>
<td>clepēh</td>
<td>RDP V.INTR AGT; me-...-an V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>60</td>
<td>cligcig</td>
<td>RDP V.INTR AGT; N- V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>61</td>
<td>dalem</td>
<td>-ang AGT, pe-...-in; AGT</td>
<td>–</td>
</tr>
<tr>
<td>62</td>
<td>darta</td>
<td>-ang AGT, pi-...-ang</td>
<td>pi- ‘explanation’</td>
</tr>
<tr>
<td>63</td>
<td>daidat</td>
<td>RDP V.INTR AGT; me- V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>64</td>
<td>deep</td>
<td>me-...-an V.INTR AGT; RDP V.INTR AGT</td>
<td>–</td>
</tr>
<tr>
<td>65</td>
<td>delēh</td>
<td>N- V.INTR AGT?; -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>66</td>
<td>delik</td>
<td>N- AGT?; -ang, -in AGT; pe-</td>
<td>–</td>
</tr>
<tr>
<td>67</td>
<td>démēl</td>
<td>N- V.INTR; -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>68</td>
<td>dēngkléng</td>
<td>N- AGT; -ang, -in AGT; RDP, pe-</td>
<td>–</td>
</tr>
<tr>
<td>69</td>
<td>dengok</td>
<td>N- V.INTR AGT?; -ang, -in AGT, RDP, pe-</td>
<td>–</td>
</tr>
<tr>
<td>70</td>
<td>derep</td>
<td>me- V.INTR AGT; -ang AGT</td>
<td>-an ‘harvested rice’</td>
</tr>
<tr>
<td>71</td>
<td>dingkrik</td>
<td>N- V.INTR AGT, RDP, pe-</td>
<td>–</td>
</tr>
<tr>
<td>72</td>
<td>dlekep</td>
<td>N- AGT; pe-, -ang AGT; RDP</td>
<td>–</td>
</tr>
<tr>
<td>73</td>
<td>dlepdep</td>
<td>N- AGT; -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>74</td>
<td>dlepēk</td>
<td>N- AGT; pe-, -ang, -in AGT</td>
<td>–</td>
</tr>
<tr>
<td>75</td>
<td>dlesdes</td>
<td>N- UG?, pe-</td>
<td>–</td>
</tr>
<tr>
<td>76</td>
<td>döngkroong</td>
<td>N-, -ang AGT</td>
<td>–</td>
</tr>
<tr>
<td>77</td>
<td>dudon</td>
<td>me-...-an V.INTR AGT</td>
<td>-an ‘order of precedence’</td>
</tr>
<tr>
<td>78</td>
<td>dagdag</td>
<td>me- UG; -ang, -in AGT, RDP</td>
<td>-an ‘pile’</td>
</tr>
<tr>
<td>79</td>
<td>dudur</td>
<td>-in AGT; me-...-an AGT</td>
<td>–</td>
</tr>
<tr>
<td>80</td>
<td>dunang</td>
<td>-in AGT; me-...-an AGT</td>
<td>-an ‘place to stay’</td>
</tr>
<tr>
<td>81</td>
<td>édēng</td>
<td>me- UG; -ang, -in AGT</td>
<td>–</td>
</tr>
</tbody>
</table>
The following items were marked as bound roots in the dictionary but were not used in counts due to insufficient evidence for their bound status:

ajag, agil, beeng, bejing, pe-cadi, cabuah, abar-abar, abih, adak, adal, adu, aeg, aguk, ajar, ajar, bancang, batek, be, bedog, bejug, beled, beles, cekeg, cékér.
Appendix 2

Sample of 200 ‘Precategorial’ Indonesian roots from Kridalaksana (1986), classified by affixing patterns. In the ‘Other’ column, \textit{ter-} and \textit{ke-} are verbal affixes, and all other affixes are nominal unless otherwise noted.

<table>
<thead>
<tr>
<th>Root</th>
<th>Meaning</th>
<th>\textit{meN-}</th>
<th>\textit{ber-}</th>
<th>\textit{Ø~an}</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 \textit{acan}_1</td>
<td>‘hope for, dream of’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2 \textit{acan}_2</td>
<td>‘make a striking motion’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3 \textit{ambul}</td>
<td>‘bounce off; fly up high’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4 \textit{ampu}</td>
<td>‘support from below’</td>
<td>✓; \textit{-kan}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5 \textit{anjur}_2</td>
<td>’pull’; -\textit{diri} ‘withdraw, resign’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6 \textit{antul}</td>
<td>‘bounce up, ricochet’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7 \textit{awur}</td>
<td>‘spread about, scatter’</td>
<td>✓ \textit{AGT}; \textit{-i}, \textit{-kan}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8 \textit{camuk}</td>
<td>1. ‘stab, whip’ 2. ‘run amok’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9 \textit{cancang}</td>
<td>‘tie up an animal’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10 \textit{cangam}</td>
<td>‘catch in one’s mouth’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11 \textit{cangkam}</td>
<td>‘hug’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>12 \textit{cangkup}</td>
<td>‘eat by throwing into one’s mouth’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>13 \textit{canguk}</td>
<td>‘sit with head down’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>14 \textit{cebik}</td>
<td>‘sneer, pushing lower lip out’</td>
<td>✓; \textit{-kan}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>15 \textit{cekau}</td>
<td>‘catch, grasp’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>16 \textit{cekit}</td>
<td>‘bite (of bird, snake). grip’</td>
<td>✓ \textit{AGT}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>17 \textit{cek}</td>
<td>‘wound by stabbing with a small, sharp object’</td>
<td>✓; \textit{-kan}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>18 \textit{cekup}</td>
<td>‘catch (e.g. a fly) in one’s hands’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>19 \textit{celang}</td>
<td>‘stare wild-eyed’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>20 \textit{celat}</td>
<td>‘jump far, high’</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>21 \textit{celis}</td>
<td>‘chop up fine’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>22 \textit{celuk}</td>
<td>‘put one’s hand into s.th. to remove s.th.’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>23 \textit{cengam}</td>
<td>‘catch in one’s mouth’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>24 \textit{cengap}</td>
<td>‘catch in one’s mouth’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>25 \textit{cengkau}</td>
<td>‘hold in claws’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>26 \textit{cengkelong}</td>
<td>‘subtract from the amount which should be given’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>27 \textit{daduh}</td>
<td>‘sing a lullaby’</td>
<td>✓; \textit{-kan}</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>28 \textit{dayu}</td>
<td>‘rock a small child in one’s arms while singing to it’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>29 \textit{depang}</td>
<td>‘spread one’s arms out to the right and left’</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>No.</td>
<td>Word</td>
<td>Meaning</td>
<td>Relevance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>-----------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>30</td>
<td>derum</td>
<td>'get down on knees (e.g. of a cow, buffalo)'</td>
<td>✓; -kan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>diar</td>
<td>'leave' (memper-...-kan)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>dongak</td>
<td>'look up'</td>
<td>✓; -kan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>dasin</td>
<td>‘wake up; come to; realise’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>embut</td>
<td>‘slowly pull a fishing line’</td>
<td>✓; RDP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>empar</td>
<td>‘move at an angle; be carried off course (boat)’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>enap</td>
<td>'settle (of sediments)'</td>
<td>✓ AGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>enjal</td>
<td>‘bounce, ricochet’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>enjal2</td>
<td>'fill one’s mouth with food'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>enjut</td>
<td>'pull slowly; paddle'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>gelosang</td>
<td>‘quickly make s.th.; rush a job’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>gelosok</td>
<td>‘rub vigorously’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>gelotak</td>
<td>'peel (fruit with a hard skin)’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>gelulur</td>
<td>'come off easily; slip; fall down (trousers, etc.)'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>gemak</td>
<td>'grip, grab at'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>gemik</td>
<td>'make a signal using the fingers'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>geragas</td>
<td>'scratch at'</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>geragau</td>
<td>'claw at’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>geragot</td>
<td>'bite at'</td>
<td>✓; -i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>geranyam</td>
<td>'move (air, because hot); feel pins and needles’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>gerapai</td>
<td>'gropo, try to grab’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>great</td>
<td>'grind teeth’</td>
<td>✓; -kan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>gerecok</td>
<td>'disturb, tempt’</td>
<td>✓; -i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>geriap</td>
<td>'make small (one’s eyes, flame of a lamp)’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>gerompok</td>
<td>'chase, surround’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>geros</td>
<td>'snore’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>gerumuk</td>
<td>'sit curled up, huddle’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>gerumut</td>
<td>'surround’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>gerundel</td>
<td>'grumble’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>gerupis</td>
<td>'do careful, fine work’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>gerupuk</td>
<td>‘fall after slipping’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>getik</td>
<td>'tickle, touch’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>getil</td>
<td>'pinch’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>gisar</td>
<td>'move, spin’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>gogok</td>
<td>'gulp down (water)’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>idam</td>
<td>‘suffer (an illness)’</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>igal</td>
<td>'spread tail feathers’ (e.g. peacock, turkey)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root</td>
<td>Meaning</td>
<td>meN-</td>
<td>ber-</td>
<td>Ø~an</td>
<td>Other</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>cangah</td>
<td>'open wide, gape'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>cangak</td>
<td>'stick head out to see'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>cekih</td>
<td>'open (shellfish, etc.)'</td>
<td>✓ AGT</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>celampak</td>
<td>'throw down violently'</td>
<td>-kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>celapak</td>
<td>'defiantly face'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>elak</td>
<td>'evade, step aside'</td>
<td>✓ -kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>empap</td>
<td>'crush, squash with heavy object'</td>
<td>✓ -kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>empas</td>
<td>'throw down' V.INTR; 'strike'</td>
<td>✓ -kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>enyak</td>
<td>'stamp on' V.TR; terper- 'suddenly fall back'</td>
<td>✓ -kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>10</td>
<td>giur</td>
<td>meng…-kan 'arouse'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>jebllos</td>
<td>'force into (a hole)'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>jegil</td>
<td>'stare wildly (eyes)'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>13</td>
<td>jerembap</td>
<td>meng…-kan 'push to make stumble'</td>
<td>✓ -kan</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>14</td>
<td>jerongkong</td>
<td>'stand bowing with hands on knees'; ter- 'fall forward to the ground'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>15</td>
<td>jeramus</td>
<td>meng…-kan 'push to make stumble'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Root</td>
<td>Meaning</td>
<td>meN-</td>
<td>ber-</td>
<td>Ø~an</td>
<td>Other</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>caran</td>
<td>‘fight verbally’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>celak</td>
<td>‘glisten, shine’</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>celatuk</td>
<td>‘chat, joke’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>cengkung</td>
<td>‘sit with one’s knees up’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>dengap</td>
<td>‘beat fast (heart)’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>dayun</td>
<td>‘arrive in great numbers’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>gembut</td>
<td>‘pulsate (like a new-born’s fontanella)’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>jingkat</td>
<td>1. ‘limp because one leg is shorter’; 2. ‘walk on tiptoes’; 3. ‘suddenly jump’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>jingkik</td>
<td>‘hop’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>10</td>
<td>jingkrah</td>
<td>‘hop, skip’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>jobak</td>
<td>‘splat; make spluttering noise while talking’</td>
<td>–</td>
<td>✓AGT</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø~an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>denyar</td>
<td>‘glisten’</td>
<td>–</td>
<td>✓AGT?</td>
<td>-an</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø~an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>agah</td>
<td>‘stare at defiantly; challenge’ (meng-, ber-)</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>anjak</td>
<td>‘move’</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>anjur</td>
<td>‘set off (on a journey)’</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>cangkung</td>
<td>‘squat’</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>cantum</td>
<td>‘bring two edges together to close’; ber- ‘close’ V.INTR (e.g. a wound)</td>
<td>✓; -kan</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>dadung</td>
<td>‘sing a lullaby to’ (meng-/ber-)</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>dampil</td>
<td>ber- ‘shut tight together’; meN-…-kan = CAUS</td>
<td>-kan</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>8</td>
<td>dekam</td>
<td>‘bow deeply’</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>dekus</td>
<td>‘purr (e.g. cat), snort (e.g. horse)’ (ber-/meN-)</td>
<td>✓AGT</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>10</td>
<td>dompak</td>
<td>1. meng- ‘rear up (horse)’ 2. ber- ‘jump when fighting’</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>érak</td>
<td>‘separate, move off’ meng- V.TR; ber- V.INTR</td>
<td>✓; kan</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>12</td>
<td>gelugut</td>
<td>‘shiver violently’ (meng-/ber-)</td>
<td>✓AGT</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>13</td>
<td>gésél</td>
<td>‘rub together’ meng- V.TR; ber- V.INTR</td>
<td>✓</td>
<td>✓AGT</td>
<td>–</td>
</tr>
<tr>
<td>14</td>
<td>genung</td>
<td>‘collect, gather (water)’ V.INTR meng-/ber-; ber- ‘flow (tears)’</td>
<td>✓</td>
<td>✓AGT+PASS</td>
<td>–</td>
</tr>
<tr>
<td>No.</td>
<td>Root</td>
<td>Meaning</td>
<td>meN-</td>
<td>ber-</td>
<td>Ø–an</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>15</td>
<td>inggul</td>
<td>'separate, move off slowly'</td>
<td>✓, -kan</td>
<td>✓ AGT</td>
<td>–</td>
</tr>
<tr>
<td>16</td>
<td>jantang</td>
<td>'be clearly visible'</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
</tr>
</tbody>
</table>

### Root Meaning
<table>
<thead>
<tr>
<th>No.</th>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø–an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>angi</td>
<td>'raise one’s head' (meng-/ber-)</td>
<td>✓</td>
<td>✓ AGT</td>
<td>–</td>
<td>ter-RDP (UV)</td>
</tr>
<tr>
<td>2</td>
<td>antuk</td>
<td>'touch, run into/bump against'</td>
<td>✓ AGT</td>
<td>✓ AGT</td>
<td>–</td>
<td>ber-, -an, ter-</td>
</tr>
<tr>
<td>3</td>
<td>jantang</td>
<td>'be clearly visible'</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>4</td>
<td>jantang</td>
<td>'suddenly meet'</td>
<td>✓</td>
<td>✓ AGT</td>
<td>–</td>
<td>ter-</td>
</tr>
</tbody>
</table>

### Root Meaning
<table>
<thead>
<tr>
<th>No.</th>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø–an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alich</td>
<td>1. meng- ‘change, replace’ 2. ber- ‘move’ V.INTR</td>
<td>✓</td>
<td>-kan</td>
<td>✓ AGT</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>arak</td>
<td>‘accompany in a crowd’ (meng/ber- )</td>
<td>✓ AGT</td>
<td>✓ AGT</td>
<td>RDP ✓</td>
<td>peng-, per..., -an</td>
</tr>
<tr>
<td>3</td>
<td>cadang</td>
<td>'be willing to, prepared to'</td>
<td>-kan</td>
<td>✓ AGT</td>
<td>✓</td>
<td>pen-, pen..., -an</td>
</tr>
<tr>
<td>4</td>
<td>diang</td>
<td>'warm oneself by a fire' (meng V.TR; ber- V.INTR)</td>
<td>✓</td>
<td>✓ AGT</td>
<td>–</td>
<td>per..., -an</td>
</tr>
<tr>
<td>5</td>
<td>édar</td>
<td>'walk around' V.INTR (ber-/meng- )</td>
<td>✓; -i, -kan</td>
<td>✓ AGT</td>
<td>✓ n</td>
<td>-er..., -an</td>
</tr>
<tr>
<td>6</td>
<td>gerayang</td>
<td>meng- ‘grasp at (e.g. of a pickpocket)’; ber..., -an 'walk about (intending to pickpocket)'</td>
<td>✓; -i</td>
<td>✓ AGT</td>
<td>-an n</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>gérék</td>
<td>meng- ‘drill a hole into’</td>
<td>✓ AGT</td>
<td>✓ RDP</td>
<td>–</td>
<td>peng-</td>
</tr>
<tr>
<td>8</td>
<td>gesa-gesa</td>
<td>'do in a rush'</td>
<td>✓; -kan</td>
<td>✓ AGT</td>
<td>–</td>
<td>ter-, ke-ter..., -an</td>
</tr>
<tr>
<td>9</td>
<td>getar</td>
<td>meng-/ber- ‘tremble’</td>
<td>✓</td>
<td>✓ AGT</td>
<td>✓</td>
<td>peng- (alat)</td>
</tr>
<tr>
<td>10</td>
<td>iring</td>
<td>ber- ‘walk in a row’; meng- ‘ accompany’</td>
<td>✓; -i</td>
<td>✓ AGT</td>
<td>✓ n</td>
<td>peng-, peng..., -an</td>
</tr>
<tr>
<td>11</td>
<td>jaja</td>
<td>ber-/meng- ‘go about hawking one’s wares’</td>
<td>✓; -kan</td>
<td>✓ AGT</td>
<td>✓ RDP</td>
<td>peN-</td>
</tr>
<tr>
<td>12</td>
<td>jalar</td>
<td>meng-/ber- ‘crawl, creep (insects, plants)’</td>
<td>✓; -kan</td>
<td>✓ AGT</td>
<td>✓</td>
<td>peN..., -an</td>
</tr>
<tr>
<td>13</td>
<td>jangkit</td>
<td>ber-/meng- ‘spread (disease)’</td>
<td>✓; -i</td>
<td>✓ AGT</td>
<td>✓ n</td>
<td>peN..., -an, ke..., -an V</td>
</tr>
<tr>
<td>14</td>
<td>jawat</td>
<td>ber- ‘hold on’; meng- ‘hold’</td>
<td>✓</td>
<td>✓ AGT</td>
<td>✓ n</td>
<td>peN-</td>
</tr>
<tr>
<td>15</td>
<td>joang</td>
<td>ber- ‘fight’; meng- ‘attack’</td>
<td>✓</td>
<td>✓ AGT</td>
<td>–</td>
<td>pe/per-, peN-per-an, ke-an</td>
</tr>
<tr>
<td>16</td>
<td>jolak</td>
<td>meN-/ber- ‘burn bright, with big flames’</td>
<td>✓</td>
<td>✓ AGT</td>
<td>✓ n</td>
<td>–</td>
</tr>
</tbody>
</table>

### Root Meaning
<table>
<thead>
<tr>
<th>No.</th>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø–an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>acu</td>
<td>'shake one’s fist or a weapon at; point to, refer to'</td>
<td>✓; -kan</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>acum</td>
<td>'stir up; challenge to fight'</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>peng-</td>
</tr>
<tr>
<td>No.</td>
<td>Root</td>
<td>Meaning</td>
<td>Categorial Status</td>
<td>Other Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>---------</td>
<td>------------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-adon</td>
<td>'mix, stir (flour, etc.)'</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>aling</td>
<td>'cover, protect'</td>
<td>✓; -i</td>
<td>-</td>
<td>✓ (n)</td>
<td>AGT</td>
</tr>
<tr>
<td>5</td>
<td>alir</td>
<td>'flow'</td>
<td>✓; -i, +-kan</td>
<td>-</td>
<td>✓ n</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>6</td>
<td>amat</td>
<td>'inspect carefully'</td>
<td>✓; -i</td>
<td>-</td>
<td>✓ n</td>
<td>peng-, peng-...-an</td>
</tr>
<tr>
<td>7</td>
<td>anggar</td>
<td>'estimate, count'</td>
<td>✓; -kan</td>
<td>-</td>
<td>✓ n</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>8</td>
<td>anggit</td>
<td>'compose'</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>anglur</td>
<td>'flow'</td>
<td>✓; -i</td>
<td>-</td>
<td>✓ n</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>10</td>
<td>antih</td>
<td>'spin thread'</td>
<td>✓ AGT</td>
<td>-</td>
<td>-</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>11</td>
<td>amat</td>
<td>'flow'</td>
<td>✓; -i</td>
<td>-</td>
<td>✓</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>12</td>
<td>awang</td>
<td>'float up high in(to) the sky'</td>
<td>✓ AGT</td>
<td>-</td>
<td>-an n</td>
<td>per-...-an, RDP v</td>
</tr>
<tr>
<td>13</td>
<td>awur</td>
<td>'do blindly'</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>RDP-an V</td>
</tr>
<tr>
<td>14</td>
<td>ayap</td>
<td>'eat/drink (food given by a ruler)'</td>
<td>✓ AGT</td>
<td>-</td>
<td>✓ n</td>
<td>ke-...-an VB</td>
</tr>
<tr>
<td>15</td>
<td>ayom</td>
<td>'protect'</td>
<td>✓; -i</td>
<td>-</td>
<td>-</td>
<td>peng-~ -an</td>
</tr>
<tr>
<td>16</td>
<td>cebak</td>
<td>'mine for'</td>
<td>✓ AGT</td>
<td>-</td>
<td>✓ n</td>
<td>penN-, pen-...-an, per-...-an</td>
</tr>
<tr>
<td>17</td>
<td>cengkam</td>
<td>'grasp'</td>
<td>✓; -kan</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>daras</td>
<td>'loudly recite the Quran'</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>pen-~ -an</td>
</tr>
<tr>
<td>19</td>
<td>ejek</td>
<td>'mock'</td>
<td>✓</td>
<td>-</td>
<td>-an n</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>embara</td>
<td>'wander about'</td>
<td>✓; -i</td>
<td>-</td>
<td>-</td>
<td>peng-, peng-...-an</td>
</tr>
<tr>
<td>21</td>
<td>embel</td>
<td>(meng-...-i) 'accompany with, add to'</td>
<td>✓; -i</td>
<td>-</td>
<td>✓ n</td>
<td>RDP n</td>
</tr>
<tr>
<td>22</td>
<td>endap</td>
<td>'settle (of sediments)'</td>
<td>✓ AGT, -kan</td>
<td>-</td>
<td>✓</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>23</td>
<td>entak</td>
<td>'stamp on' V.TR; 'throb (e.g. heart)' V.INTR</td>
<td>✓; -kan</td>
<td>-</td>
<td>-</td>
<td>peng-</td>
</tr>
<tr>
<td>24</td>
<td>geropyok</td>
<td>'(of many people) catch (e.g. a thief)'</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>peng-~ -an</td>
</tr>
<tr>
<td>25</td>
<td>godok</td>
<td>'boil' V.TR</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>peng-an</td>
</tr>
<tr>
<td>26</td>
<td>idap</td>
<td>'suffer from a chronic disease'</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>peng-</td>
</tr>
<tr>
<td>27</td>
<td>igau</td>
<td>'suffer from a chronic disease'</td>
<td>✓</td>
<td>-</td>
<td>✓ n, RDP V.INTR</td>
<td>peng-</td>
</tr>
<tr>
<td>28</td>
<td>imbau</td>
<td>'call on, make a plea for' V.TR</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>ter-</td>
</tr>
<tr>
<td>29</td>
<td>impi</td>
<td>'dream'; -kan 'dream of, hope for' V.TR</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>ter-</td>
</tr>
<tr>
<td>30</td>
<td>inap</td>
<td>'spend the night' V.INTR</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>31</td>
<td>ingar</td>
<td>RDP 'very noisy, be a racket'</td>
<td>✓; -kan</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>ingau</td>
<td>'talk involuntarily (e.g. in one’s sleep)'</td>
<td>✓</td>
<td>-</td>
<td>✓ ADJ</td>
<td>-</td>
</tr>
<tr>
<td>33</td>
<td>ipuk</td>
<td>'scatter (seeds)'</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>peng-...-an</td>
</tr>
<tr>
<td>34</td>
<td>iras</td>
<td>'become one'</td>
<td>✓</td>
<td>-</td>
<td>✓ n</td>
<td>-</td>
</tr>
<tr>
<td>Root</td>
<td>Meaning</td>
<td>meN-</td>
<td>ber-</td>
<td>Ø~an</td>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>irik</td>
<td>'tread on'</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>peng-</td>
<td></td>
</tr>
<tr>
<td>jabar</td>
<td>'explain in detail'</td>
<td>✓, -kan</td>
<td>–</td>
<td>✓ n</td>
<td>ter-…-kan; peng…-an</td>
<td></td>
</tr>
<tr>
<td>jajak</td>
<td>men-…-i 'analyse; suspect'</td>
<td>✓, -i</td>
<td>–</td>
<td>–</td>
<td>peng-an</td>
<td></td>
</tr>
</tbody>
</table>

Roots with apparently non-agentive subjects: 8 (4%)

<table>
<thead>
<tr>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø~an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>angah</td>
<td>'be speechless'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>cengung</td>
<td>'be motionless, speechless with astonishment'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>engah</td>
<td>'gasp for breath'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>geluncur</td>
<td>'slip, slide'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>gema</td>
<td>'be speechless, startled and confused'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>jajak</td>
<td>men-…-i 'analyse; suspect'</td>
<td>✓, -i</td>
<td>–</td>
<td>–</td>
<td>peng-an</td>
</tr>
<tr>
<td>jila</td>
<td>'lie sprawled on the ground'</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>ter-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Root</th>
<th>Meaning</th>
<th>meN-</th>
<th>ber-</th>
<th>Ø~an</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>anja-anja</td>
<td>'behave in a spoiled manner'</td>
<td>–</td>
<td>✓ AGT</td>
<td>–</td>
<td>ter- (UV)</td>
</tr>
<tr>
<td>cengang</td>
<td>'startled, astonished'</td>
<td>-kan</td>
<td>✓; -an PAT</td>
<td>–</td>
<td>ter-</td>
</tr>
<tr>
<td>gelumang</td>
<td>'dirty and muddy, covered in mud'</td>
<td>–</td>
<td>PAT</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>jebai</td>
<td>'in a disorderly mess'</td>
<td>–</td>
<td>✓ PAT</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>jerumut</td>
<td>'sew by hand'</td>
<td>✓</td>
<td>✓ PAT?</td>
<td>–</td>
<td>peN-</td>
</tr>
</tbody>
</table>
The items below were not counted in totals:

1. two forms which surfaced with -an only, both loans from Javanese, hence they can be regarded as monomorphemic: *imbalan, aleman*.
2. forms considered erroneously listed as bound by Kridalaksana — e.g. where the KBBI listed them as free morphemes, or I independently found a usage as a free morpheme.
3. forms listed by Kridalaksana which were not in my *Kamus Besar Bahasa Indonesia* (KBBI).

<table>
<thead>
<tr>
<th>Root</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>aleman</td>
<td>Javanese loan, monomorphemic in Indonesian</td>
</tr>
<tr>
<td>imbalan</td>
<td>Javanese loan, monomorphemic in Indonesian</td>
</tr>
<tr>
<td>andal</td>
<td>free</td>
</tr>
<tr>
<td>asa</td>
<td>free</td>
</tr>
<tr>
<td>bugar</td>
<td>free</td>
</tr>
<tr>
<td>canda</td>
<td>free</td>
</tr>
<tr>
<td>candan</td>
<td><em>ke-candan</em> Javanese loan, <em>canda</em> free</td>
</tr>
<tr>
<td>cawang</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>desas</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>gemanl</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>gerutu</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>imbang</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>inas</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>indang</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>inggang</td>
<td>reduplicated variant of free morpheme <em>inggaung</em></td>
</tr>
<tr>
<td>isak</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>jangkang</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>jerkat</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>jerumbai</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>jilah</td>
<td>free in KBBI</td>
</tr>
<tr>
<td>dansi</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>darau</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>dedal</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>dekan</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>dengkik</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>dikit</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>gerondang</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>gocok</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>jaksi</td>
<td>not in KBBI</td>
</tr>
<tr>
<td>jingau</td>
<td>not in KBBI</td>
</tr>
</tbody>
</table>
References


—— 1998, Symmetrical voice systems and precategoriality in Philippine Languages.


1 Introduction

A comprehensive framework for discussing demonstratives has been provided by recent analyses of their functions (Himmelmann 1996; Dixon 2003). The present paper describes the morpho-syntax of the Tokelauan demonstratives, and uses the taxonomy provided by Himmelmann to discuss the uses of the nominal demonstratives in a range of texts, with the additional focus of determining how the various functions are distributed across text types. It will be shown that discourse deictic uses predominate in all text types except folktales, and are more frequent in proportion to the degree of affect or opinion expressed by a text, and to the skills of the speaker in shaping a text. Demonstratives give structure to a text, particularly at the level of the sentence or the thematic paragraph. In particular, focus constructions containing demonstratives are a common rhetorical means of indicating the point of a story or of asserting or reinforcing an argument.

The corpus of approximately 55,000 words used for this study consists of recorded and transcribed Tokelauan collected between 1967 and 1986. It includes the following types of text: personal narratives, folk tales, expository texts about fishing techniques, a meeting of the village council, a sermon, and a formal public speech. It does not include informal conversational data, apart from representations of dialogue in narratives, but the village meeting has many of the characteristics of conversation, a fact which is reflected in the kind of demonstrative uses which occur. The texts in the corpus are discussed individually in §5.1

2 The Tokelauan demonstratives

Two demonstrative particles, tē ‘this, that’ (transparently related to the singular specific determiner te) and iē ‘these, those’, are frequently combined with the following deictic particles, which in their situational use indicate position in relation to the participants in a conversation: -nei (near speaker), -nā (near addressee), and -ia (away from both speaker and addressee). Tē and iē are used without the position particles, typically in gestural

1 It gives me much pleasure to dedicate this paper to Andy Pawley, a friend, teacher and mentor of long standing. My first published paper was written at his instigation and concerned an unpromising-looking Oceanic particle. The experience was fruitful, and gave me an abiding interest in the semantics and pragmatics of grammatical forms, which I am here pursuing once again.
deixis. In addition, *nei* occurs on its own with the same meaning as *tēnei* and *iēnei*, and *ia*\(^2\) also occurs on its own. The full paradigm of nominal demonstratives is given in Table 1.

**Table 1**

<table>
<thead>
<tr>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>near speaker/proximal/’what I am about to say’</td>
<td><em>tēnei, nei</em></td>
</tr>
<tr>
<td>near addressee/distal/’what I just told you’</td>
<td><em>tēnā</em></td>
</tr>
<tr>
<td>‘yonder’/contrastive/’what I mentioned earlier’, ‘what we all know’</td>
<td><em>tēia, ia</em></td>
</tr>
</tbody>
</table>

These forms are used adnominally and pronominally. The proximal demonstratives are used frequently for temporal deixis, and the postverbal particle *nei* ‘now’ would appear to be derived from the proximal demonstrative.

There are other demonstrative forms. The deictic particles -*nei* and -*nā* are suffixed to the preposition *ki* to form the locative nouns *kinei* ‘here, this place’ and *kinā* ‘there, that place’. The three particles also occur suffixed to the verb *vē* ‘be like’ to form the deictic verbs *vēnei* ‘be like this’, *vēnā* ‘be like that’, and *vēia* ‘be like that’. Tokelauan thus has the three ‘well-attested’ types of demonstrative listed by Dixon (2003:62), nominal, local adverbial and verbal, with the proviso that the locative nouns must be preceded by a preposition to construct a local adverbial phrase. An additional feature is that the deictic verbs can occur as post-modifiers of other verbs, thus functioning as manner adverbial demonstratives: *fai vēnā* ‘do like that’; and as post-modifiers of nouns: *ni lapa vēnā* ‘coral heads of that kind’. Tokelauan thus provides an exception to Dixon’s generalisation that such demonstratives are ‘almost always morphologically derived from nominal demonstratives’ (2003:62). Only the nominal demonstratives are discussed in this paper.

3 The syntax of Tokelauan nominal demonstratives

The features of Tokelauan syntax that are most relevant here are the structure of the noun phrase and the structure of nominal clauses.

3.1 Demonstratives used adnominally

The constituents of nominal phrases with a common noun as nucleus have the following typical order of occurrence (optional items are enclosed in brackets):

(preposition)-determiner-(possessive pronoun)-(premodifier)-**nucleus**-(postmodifier)-(demonstrative).

The nucleus consists of a single lexeme, such as *fale* ‘house’, or of a head lexeme and one or more lexemes used as qualifiers. As this schema shows, adnominal demonstratives are the last item to occur in a noun phrase, apart from relative clauses and possessive phrases.

\(^2\) Not to be confused with its several homonyms: third person singular pronoun, absolutive marker, and discourse particle.
Demonstratives in Tokelauan discourse

(1) nā puha nei ē nei
SPEC.PL box DEM
‘these boxes’

(2) Ko ai te tino teiia?
PR who SPEC person DEM
‘Who is that man over there?’

(3) nā hāhā muamua iēnā o te hāuai
SPEC.PL mystery first DEM.PL POSS.INAL SPEC ogre
‘those first mysteries of the ogre’ (TTT:74)

In addition, demonstratives can premodify a noun. In this case they replace the determiner.

(4) Mai te taimi na fano ai tēnei tama, e hēki mapu
from SPEC time TA go ANAPH DEM youth TA NEG rest
tēnei toeaina.
DEM elder
‘From the time this young man went away, this old man had no repose.’ (S:37)

They can modify pronouns, although this is not common:

(5) Ko au tēnei e kō faia.
PR 1s DEM TA 1s do
‘It is I here who did it.’ (TTT:64)

(6) Ke manatua e kitātou uma, kitātou iēnē e nonofo
MOD remember ERG 1pi all 1pi DEM TA sit:PL
i te mulivaka, ko tagata uma lele i tō
LOC SPEC rear.canoe PR man all INTSF LOC 2s.POSS.INAL
fāoa, ko ni tautai uma lava kilatou iēnā.
crew PR NSPEC.PL fisher all INTSF 3p DEM
‘Let us all remember, those [lit. these] of us who sit at the stern, that all
the men in your crew, those ones are all master-fishermen too.’ (PTT:12)

3.2 Demonstratives as the nucleus of a noun phrase

Pronominal demonstratives occur most frequently in a nominal clause, as either the
predicate or the absolutive constituent. These structures will be discussed in the next
section.

I have found a few examples of a demonstrative functioning as an absolutive in a verbal
clause. In (7) the pronominal status of tēnei is made clear by the absolutive marker ia.

(7) Ia, olo uma foki nā tino kae tū lava
D.PTCL go:PL all INTSF SPEC.PL person CNJ stand INTSF
ia tēnei, a!
ABS DEM D.PTCL
‘Yes, all the people went off, but here am I just standing eh!’ (PN1:13)
(8) E olo fakatahi kimātou oi hakili mai nā mea ika,
TA go:PL all.together 1pe SQ seek DIR SPEC.PL thing fish
kaumai ai ki te matai o te kaiga e felau
bring ANAPH to SPEC headman POSS.INAL SPEC family TA share.out
ai, ko tēnā e takua ai he utuga e
ANAPH PR DEM TA be.known.as ANAPH NSPEC division.of.fish TA
olo kimātou ki ei.
go:PL 1pe to ANAPH
'We go all together and harvest fish, bring it back to the matai of the extended family who shares it out, and that is known as a [formal] division of fish which we go to.' (WF:235)

A pronominal demonstrative can occur in construction with the locative preposition i, as in (9). In (10) the fronted tēnā is represented in the clause by the oblique anaphor ai, suggesting some kind of relationship to a phrase of the form i tēnā.

(9) E hē gata i tēnā te tūlaga a te tamāloa.
TA NEG stop LOC DEM SPEC behaviour POSS.AL SPEC man
'The fellow’s behaviour didn’t stop at that.’ (PN2:2)

(10) Ko tēnā lā e poto ai ia Matulafoaina i
PR DEM INTSF TA know ANAPH ABS Matulafoaina LOC
nā uiga o te hāuai.
SPEC.PL habit POSS.INAL SPEC ogre
'It was because of that [a magic object] that Matulafoaina was knowledgeable in the ways of the ogre.’ (TTT:72)

The demonstratives tēnei and tēnā occur frequently at the beginning of a clause, functioning rather like sentence adverbs or free topics. In this use, tēnā is translatable as ‘thereupon’ or ‘thus’, and makes comprehensive reference to a set of circumstances that has been adumbrated in the preceding context, thus constituting an example of discourse deixis. This construction is distinguished from examples like (10) above where the fronted demonstrative is firmly anchored in the structure of the following clause.

(11) Tēnā lava, kamata fakaputuputu heleni.
DEM INTSF begin CAUS:be.close.together shilling
'Things being that way, I began to save money.’ (PN1:2)

When tēnei occurs in the same position it usually has temporal significance, ‘at this point in the narrative’:

(12) Kae tēnei kua pupuna nā vai vevela a Malililelautoa
CNJ DEM TA boil SPEC.PL water hot POSS.AL Malililelautoa
e fai ai nā tamaiti.
TA do ANAPH SPEC.PL child
'But now Malililelautoa’s hot water is on the boil, for cooking the children.’ 
(TTT:56)

3 Compare Mosel (2004:152, §5.4) on demonstratives functioning as sentence-initial adjuncts.
3.3 Nominal clauses

Nominal clauses differ from all other kinds of independent clause in that they do not contain a tense-aspect particle. They can consist of a nominal predicate alone, which has a presentative or identifying function and is introduced by the presentative particle ko. Otherwise they contain two noun phrases, the predicate and the unmarked absolutive, constituting an equational clause. The predicate may identify the referent of the absolutive noun phrase, or it may assign it to a class. When the predicate is indefinite, as in the second example below, it always signifies membership in a class. Indefinite predicates do not have to be preceded by ko, and in fact ko is often omitted under other circumstances, and particularly if the predicate consists of a demonstrative, see (15) below. Both absolutive and predicate NPs may be modified by possessive phrases or relative clauses.

(13) Ko tana gāluega tēnā.
PR 3s.POSS.AL work DEM (predicate) (absolutive)
‘That is his work.’

(14) He mea matagā tēnā.
NSPEC thing ugly DEM
‘That is an ugly state of affairs.’ (PTF:13)

(15) Tēnā te mea e māfua ai toku tāofi.
DEM SPEC thing TA be.caused.by ANAPH 1s.POSS.INAL opinion
‘That is the thing which is the source of my opinion.’ (PTT:133)

(16) Tēnei kō lā to mā fitā e fai.
DEM INTSF INTSF POSS.INAL 1de tiredness TA do
‘This very person here is the cause of our vexation.’ (TTT:64)

Modifiers of the predicate NP are always extraposed if the absolutive constituent is a demonstrative, as in (17).

(17) Ko te ala ia o toku fia hola.
PR SPEC reason DEM POSS.INAL 1s.SPEC.POSS.INAL wish escape
‘That was the reason for my wishing to run away.’ (PN2:4)

(18) Ko te togafti tēnei kua kō mautinao kua hili atu
PR SPEC remedy DEM TA 1s certain TA be.best DIR
tē lelei
SPEC good
‘The following is the remedy that I am certain is the most effective.’ (PTF:34)

Tokelauan has basic predicate-initial word order, but the fronting of a noun phrase is a common word-order permutation which allows any NP of a clause to occur clause initially, preceded by ko. Oblique arguments which appear in this position are represented anaphorically in the clause, as in (10) above. Such ko-phrases have several discourse functions related to topicalisation or focussing. Quite distinct from that construction, a sentence-initial NP introduced by ko can function as a free topic:
Ko-fronting can apply to the absolutive NP of a nominal clause. In the canonical case there will thus be two NPs introduced by ko, the first being the fronted absolutive and the second the predicate. The following passage is not unusual in its proliferation of ko-phrases in the different functions described:

(19) Ko te fāpuku, tēnā lava te mea e tukufua ai.
PR SPEC rock.cod DEM INTSF SPEC thing TA spawn ANAPH
‘As for the rock cod, that indeed is the area where it spawns.’ (PTT:136)

Ko-fronting can apply to the absolutive NP of a nominal clause. In the canonical case there will thus be two NPs introduced by ko, the first being the fronted absolutive and the second the predicate. The following passage is not unusual in its proliferation of ko-phrases in the different functions described:

(20) Ko te toe nofoaga lava o te tino,
PR SPEC final seat INTSF POSS.INAL SPEC person
fronted absolutive
ko te foemua.
PR SPEC bow-paddler
predicate
Ko te tino lā tēnā i mua,
PR SPEC person INTSF DEM LOC front
free topic
ko te faiakoga fehoahoani tēnā o te tautai.
PR SPEC teacher helper DEM POSS.INAL SPEC fisher
predicate absolutive extraposed possessive
The very final position of the crew member is [as] the bow paddler. That man in the bow, that one is the master fisher’s assistant teacher.’ (PTT:26)

(21) Ko te togafiti tēnā, he mea kua kō mauninoa.
PR SPEC remedy DEM NSPEC thing TA is certain
fronted absolutive predicate
‘That remedy is a thing about which I am quite certain.’ (PTF:39)

My corpus abounds in sentences which start with a ko-marked NP, followed by a demonstrative, followed by a verbal clause. These are structurally ambiguous between (a) a verbal clause with a ko-fronted NP that is postmodified by a demonstrative, and (b) a nominal clause with a demonstrative as the absolutive constituent, and an extraposed relative clause on the predicate NP. The rules for anaphora within the clause are the same in both cases.

(22) Ko te aholua tēnā e tatanu ai te ugauga.
PR SPEC right.day DEM TA burrow ANAPH SPEC coconut.crab
a. ‘In that season the coconut crab goes into its burrow.’

b. ‘That is the season when the coconut crab goes into its burrow.’ (PTT:84)

This structure, with the (b) interpretation, is especially prone to rhetorical uses which will be discussed later. In the absence of other cues, commonsense intuitions about the information structure of the passage usually clear up any semantic ambiguity.

However, another point needs to be made about information structure. It is assumed that in the unmarked case a predicate contains the ‘new’ information. In a nominal sentence with a fronted absolutive, the fronted constituent gains prominence, and thus both constituents are in some sense in focus. Moreover, when neither element is marked with ko, one must rely on context to determine which is the predicate and which the absolutive.
However, I don’t think this often matters. Nominal clauses containing demonstratives do not always introduce new information — often they reinforce an argument. In the following example, the function of the predicate, which I am assuming to be the demonstrative, is to throw emphasis upon the referent of the absolutive NP, with a summing-up function. See also (19) above, where the intensifier lava helps to identify tēnā as the predicate.

(23) Tēnā te uiga o toku faufautuaga mōni.

DEM SPEC meaning POSSINAL 1s.POSSINAL advice true

‘That was the reason for my sincere advice.’ (PN2:6)

The all-purpose demonstrative for this and other rhetorical uses is tēnā.

3.4 Frequencies

In all texts the singular demonstratives are overwhelmingly more frequent than the plural ones, partly reflecting the greater proportion of singular referents and partly because discourse deixis usually involves singular demonstratives. The least used are tēia and iēia. More noteworthy is the great frequency of tēnā. In the Perez texts there are substantially more occurrences of tēnā than of all the other demonstratives combined. Surprisingly, in the much shorter ‘Tale of Tautilifaga’, an example of a completely different genre, very similar proportions are found. There are 36 tokens of tēnā as opposed to 31 tokens of all other demonstratives.

4 The meanings and functions of nominal demonstratives

In this section I discuss briefly the classification of discourse uses of demonstratives in Himmelmann (1996) and apply it to the Tokelauan data, with some modifications suggested by the nature of these texts. In particular, it must be stressed that the boundaries between categories are quite porous, and that some tokens of demonstratives appear to fulfill more than one function. There is no space here to discuss these complexities. The reader is referred to Himmelmann’s comments on ‘overlaps and transitional areas’ (1996:218–219, 240–242).

4.1 Situational use

In its clearest manifestation, situational use involves reference to an entity present in the context of utterance, or to a time which is related to the time of utterance. This use accounts for the canonical definitions of demonstratives in descriptive grammars (and also in Table 1 above), but actually plays a very minor part in all but two of the texts examined here. It would probably figure more prominently in conversational data — though an examination of reported dialogue in the folktales does not provide many examples.

(24) Ko te Aitufolopuga tēnei e lea kia Tokalalaga, “Pepelo PR SPEC Ogre.swallow.coral DEM TA say to Tokalalaga lie kō  lā, ko au tēnei ia Hina, ko te Aitufolopuga tēnā.” INTSF PR 1s DEM ABS Hina PR SPEC Aitufolopuga DEM ‘The Aitufolopuga here [the Ogre-who-swallows-coral] says to Tokalalaga, “You are lying, it is I here who is Hina, that one is the Aitufolopuga”.’ (TTT:16)
The highest proportion of situational uses occurs in the Village Council Meeting and the Formal Speech. These are highly situated texts both locally and temporally.

(25) I te täea nei, e momoli atu ai ta mātou LOC SPEC morning DEM TA convey DIR ANAPH POSS.AL 1pe fakamālō ma te fakafetai ki Mālō Kaufakatahi ona compliment and SPEC thanks to government group.all.together because ko toutou amanakia mai o tēnei Atunuku vāivai. PR 2p.POSS.INAL take.notice DIR POSS.INAL DEM country weak ‘On this morning we convey our compliments and gratitude to the United Nations for taking notice of this powerless country.’ (UN:5)

4.2 Reference tracking

The reference tracking use is an emphatic form of mainly anaphoric reference to participants in a narrative. It occurs in the folktales and the autobiographical narratives, but is not a major use elsewhere. I also find a number of tokens where the entity in question is inanimate, and not in any way a major participant, but short of creating another function entirely, I have considered these a type of tracking use, see (29) and (34) below. The demonstrative typically used for anaphoric tracking of all participants is tēnā, thus differing from Samoan where the proximal demonstrative is used for thematically prominent participants (Mosel 2004:161). However, we shall see that there are two contexts in which tēnei is used for major participants.

Since zero anaphora is a common form of reference tracking of major participants in narrative, especially in the case of ergative constituents (Hooper 2000:167–70), the question must be addressed as to why demonstratives are used at all for tracking. The answer, unsurprisingly, is for emphasis and contrast. It is also worth noting that a significant number of tracking uses of demonstratives occur in oblique NPs, where ellipsis is not a possibility. The following uses and contexts are observed.

4.2.1 Highlighting major participants

Nominal demonstratives are used to highlight major participants, especially just after the first mention (cf. Lichtenberk 1996:384ff.). The almost formulaic pattern described by Mosel for Samoan (2004:160–161) occurs in Tokelauan folktales also, although not so pervasively.

(26) Nonofo, nonofo te ulugālī, fai ta lā tama, he tama live:PL SPEC couple make POSS.AL 3d child NSPEC child teine. Ko te tama ulumatua tēnā o to lā nonofoga. girl PR SPEC child firstborn DEM of POSS.INAL 3d live:PL.NR ‘Living and living together the couple produced their child, a female child. That was the eldest child of their union.’ (TT:92)

In the Tokelau narratives this first anaphora does not always occur in the clause immediately following the first mention of the participant. It always involves tēnā, not the proximal form as in Samoan, and the pattern is sometimes repeated through several
subsequent mentions. In what seems a similar function of establishing a context, tēnā can modify a non-human noun in its first mention, if the referent of the noun is somehow pragmatically presupposed in the context.

(27) Ko Lū ma Lekava he ulugali. Ko Lū te tagata, ka ko Lekava PR Lu and Lekava NSPEC couple PR Lu SPEC man CNJ PR Lekava
    te fafine. Ko te māhani lā o te kautagata SPEC woman PR SPEC custom INTSF POSS.INAL SPEC population
    o te fenua tēnā …
    POSS.INAL SPEC land DEM
    ‘Lu and Lekava were a couple. Lu was the husband and Lekava was the wife.
    Well, the custom of the people of that land …’ (TT:22)

4.2.2 Returning to a prominent narrative participant after a break

In what I will call the ‘meanwhile’ function, tēnei is used in a distinctive construction in which the NP referring to the participant is fronted, preceded by the contrastive conjunction ka/kae, and the narrative-advancing information follows.

(28) Ka ko Tautilifaga tēnei kua fanaifo ki gātai.
    CNJ PR Tautilifaga DEM TA go:DIR to shore
    ‘But meanwhile here is Tautilifaga who is going down to the sea.’ (TTT:70)

4.2.3 Contrast

The same construction can be used for thematic contrast. The following example comes from an explanatory tale in which different parts of the tropic bird are consumed by various fishes, which consequently have different feeding characteristics.

(29) Ia, ka ko te kanofi lā tēnei o te Tavake,
    D.PTCL CNJ PR SPEC flesh INTSF DEM POSS.INAL SPEC tropic.bird
    kua kai e ika o te uluulu.
    TA eat ERG fish POSS.INAL SPEC bait.fishing
    ‘Well, the flesh here of the Tropic Bird, it was eaten by fish that come to bait.’
    (TT:62)

4.2.4 First-person reference

Tēnei is used to highlight a first-person reference in a narrative. The examples are all from PN1 and occur at places in the narrative where a level of affect and concern is directed towards the predicament of the narrator (see also example (7) in §3.2).

(30) Kua uma nā kaiga a nā tino ka ko tēnei
    TA finish SPEC.PL eat:NR POSS.AL SPEC.PL person CNJ PR DEM
    lava au koi tatipiti.
    INTSF 1s TA cut:PL.RDP
    ‘The [other] people have finished eating but here am I still just cutting away.’ (PN1:8)
4.2.5 Emphasis

The emphatic use is achieved usually by means of a fronted NP modified by a demonstrative, or in a nominal clause with a demonstrative absolutive.

(31) Ko te mea tēnā e olo koutou ki ei, e hēai lele
PR SPEC thing DEM TA go:PL 2p to ANAPH TA NEG INTSF
he malaga e Hao ai.
NSPEC journey TA escape ANAPH
‘As for that place you are going to, no travellers ever escape from there.’ (TTT:66)

4.2.6 Hypothetical or generic reference

A variant of the tracking use is pervasive in procedural discourse, in which hypothetical or generic entities and persons are spoken about. There are examples in (20) above, and in the following translated passage.

(32) ‘There is a very dubious practice of a certain sort of fellow, like this: when he gets to a good coral head and the cod are biting well, that fellow (te tino tēnā) just leaves. He moves off and settles down further on. Then those people (nā tino iēia) [mentioned earlier], who have just reached that coral head (te lapa tēnā), think that the reason that man (te tino tēnā) has settled down further on, its meaning is that he has indeed found a [good] coral head. But no. He is lurking there so that you will go away. And so your group goes on and on, and then that man (te tino tēnā) returns to the previous coral head.’ (PTT:126)

4.3 Recognitional use

The recognitional use occurs occasionally in my data and has the features found in other languages, that is to say it draws on shared knowledge, often external to the context, presented in a relative clause or other modifier, to identify a new referent of low salience in the discourse context. It can also use previously introduced information to remind the hearer about a recurrent but not highly salient referent.

(33) Kē iloa ia Pihi, tēia tino Elihe?
2s know ABS Pihi DEM person Ellicean
‘Do you know Pihi, that Ellicean fellow?’ [PN1:10]

(34) Kae poto lele ia Matulaoaina, auā foki lā nā
CNJ clever very ABS Matulaoaina because INTSF INTSF SPEC.PL
mea iēnā na foki ki ei e tona tamana.
thing DEM TA give to ANAPH ERG 3s.POSS.INAL father
‘But Matulaoaina understood very well, because of those things that had been given to him by his father.’ (TTT:72)

4.4 Textual deixis

Lyons (1977:667–668) makes a distinction between ‘pure textual deixis’, in which the referents are linguistic entities (e.g. lexemes, sentences, parts of a narrative), and ‘impure
textual deixis’, in which the referents are propositions, facts and states of affairs. The latter is an unfortunate expression for what is a widespread and important use of the demonstratives, and I will substitute the term ‘discourse deixis’, see §4.5. I use ‘textual deixis’ only when the referent is a linguistic entity.

However, note that the term ‘textual deixis’ relates simply to the type of referent, and examples may have different discourse functions. So (35) could also be classed as situational deixis. Otherwise examples tend to be quite straightforward, with proximal forms used cataphorically and distal forms anaphorically.

(35) *Ia. Ko te tamā kakai ia.*
D.PTCL PR SPEC small tale DEM
‘Well, That was the short story.’ (TTT:38)

(36) *E fakamolemole atu ki te kau fakalogologo, toe fakafoki mai* TA apologise DIR to SPEC group listen:PL again retract DIR *te kupu tēia ‘nūtia.’ E hehē!* Kae ko te kupu hako tēnei: SPEC word DEM break TA mistake CNJ PR SPEC word correct DEM
‘E hehē lōtia mea. He ika e mou-laga.’ Ko te kupu TA NEG ? thing NSPEC fish TA disappear-rise PR SPEC word
hako tēnā.
correct DEM
‘I apologise to my group of listeners, I withdraw that word ‘nūtia’. It was a mistake! But this is the correct expression: “It is not a lōtia situation. The fish have stopped rising.” That is the correct expression.’ (PTF:56)

In example (36), *tēia* indicates a referent that was mentioned some while back. However *tēia* is often used instead of *tēnei* for cataphoric reference to a linguistic element assumed to be general knowledge.

(37) *E i ei te kupu tēia: kua ‘tokia’ te pala.* TA LOC ANAPH SPEC word DEM TA be.struck SPEC wahoo
‘There is that expression [that we know about]: the wahoo is “scared off”.’ (NW:278)

In the following passage both *tēnei* and *tēia* are used cataphorically, with the function of *tēia* being apparently to point out a contrast with the previous phrase.

(38) … *e uīgā ki te kupu tēnei ko te ‘tau ki mua’* TA be.about to SPEC word DEM PR SPEC count to front
ma te kupu tēia ‘tau ki muli’ (te mahina).
and SPEC word DEM count to back SPEC moon
‘… concerning this term “forward count” and that other term “backward count” (of the moon).’ (PTT:97)

4.5 Discourse deixis

Situational, reference tracking, recognitional and textual deixis are the bread and butter functions of demonstratives. It is predominantly in discourse deixis that demonstratives are
used to add emphasis, flavour and affect to a text. Both adnominal and pronominal demonstratives are used for discourse deixis, and this is by far the commonest use of the pronominal forms. In Tokelauan, as in many other languages (Himmelmann 1996:225), demonstratives appear to be the only means of expressing discourse deixis.

(39) Tēnā lā e fakamāonia ai.
DEM INTSF TA be.proved ANAPH
‘What I have just described has been attested.’ (NW:155)

Recall that in discourse deixis the referents are propositions and states of affairs. Himmelmann points out that ‘no referent exists in advance to which one may point. Instead, the referent is first created at the very moment when this use occurs’ (1996:224). Consider (8) above, in which a complex series of events is given a name, and also the following.

(40) Kāfai ko tona mātua e i to mātou kāiga, ka ko
if PR 3s.POSS.INAL mother TA LOC POSS.INAL 1pe family CNJ PR
te itu ki tona tamana, e lahi o lātou fenua, ia,
SPEC side to 3s.POSS.INAL father TA large POSS.INAL 3p land D.PTCL
e i ei te advantage tēnā e maua mai iate kilātou.
TA LOC ANAPH SPEC advantage DEM TA obtain DIR LOC 3p
‘If his mother is in our extended family, but on the side of his father they have a lot of land, well, there is that advantage that accrues to them.’ (PN1:3)

In (40), a number of other NPs could have been used instead of ‘te advantage tēnā’ to characterise the situation that has been described: ‘disadvantage’ (from the speaker’s point of view), ‘unfairness’, ‘injustice’, and so on. Only when the word ‘advantage’ is used is the situation characterised as such.

In addition to propositions, facts and states, Himmelmann includes in this category ‘reference to a point in time in a sequence of narrated events’ (1996:225), and I would want to add reference to periods of time, such as ‘in those days’. These are distinct from temporal references which are situational and are closely anchored to the speech event, such as ‘today’ and ‘yesterday, or ‘in these days’. Procedural texts provide good instances of temporal discourse deixis.

(41) Kafai e heheu mai, oho, oho, oho, pā ai ki te
when TA draw.the.lure DIR leap arrive ANAPH to SPEC
taimi tēia ka liliu ai te pā ka toe hahau,
time DEM TA turn ANAPH SPEC lure TA again cast
ko te taimi tēnā kua oho ai te atu.
PR SPEC time DEM TA leap ANAPH SPEC skipjack
‘When you are skipping the lure in there will be continual rises, [until] it comes to that time [which we know about] when the lure changes direction as you are about to make another cast, it’s at that time [just mentioned] that the skipjack bites.’ (PTF:111)

In all the Tokelauan texts under consideration except the folktales, discourse deixis is the most frequent function of the demonstratives.
5 Demonstratives and discourse types

In this section I describe the texts in the corpus and discuss the different ways demonstratives are used in them. The frequency of demonstratives in each text and the percentages of different functions are displayed in Table 2.

<table>
<thead>
<tr>
<th>Words</th>
<th>Demonstratives/ Situational</th>
<th>Discourse</th>
<th>Tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Narrative¹ (PN1)</td>
<td>11,330</td>
<td>1/79</td>
<td>–</td>
</tr>
<tr>
<td>Personal Narrative² (PN2)</td>
<td>3,360</td>
<td>1/100</td>
<td>–</td>
</tr>
<tr>
<td>Tale of Tautifulaga (TTT:62–84)</td>
<td>4,720</td>
<td>1/88</td>
<td>–</td>
</tr>
<tr>
<td>Sermon (S)</td>
<td>2,240</td>
<td>1/83</td>
<td>–</td>
</tr>
<tr>
<td>UN Speech (UN)</td>
<td>526</td>
<td>1/83</td>
<td>–</td>
</tr>
<tr>
<td>Village Council (VC)</td>
<td>5,000</td>
<td>1/38</td>
<td>–</td>
</tr>
<tr>
<td>Peato Perez (PTF), (PTT)</td>
<td>16,000</td>
<td>1/43</td>
<td>–</td>
</tr>
<tr>
<td>Noosing Wahoo (NW)</td>
<td>6,410</td>
<td>1/92</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: This Table shows the number of words in each text, the proportion of demonstratives to other words, and the percentages of the main functions in each text. A gap in the Table indicates an insignificant use of that function. No columns have been assigned to textual or recognitional uses as these are of low frequency in all cases.

5.1 Personal Narrative 1 (PN1)

In Personal Narrative 1 (PN1) a Tokelauan man who immigrated to New Zealand in his early twenties is interviewed about 15 years later by another, roughly contemporary, Tokelauan male. The speaker is unused to this kind of performance and there are many incomplete or repaired utterances and hesitations, indicative of the unplanned nature of the talk. There are more than twice as many discourse uses as all the others combined, including numerous phrases of the form: Ko tota māfaufau tēnā (PR 1s.EMP.POSS.INAL) thought DEM ‘That is my opinion ….’ (PN1:25). The speaker is emotionally involved in his tale and concerned to present a particular image of himself.

(42) E tatau ona fai te mea tēia e kō iloa, e work ni!
TA necessary CMP do SPEC thing DEM TA 1s know TA work D.PTCL
‘I had to do that thing that I knew how to do: work, eh!’ (PN1:18)

As we saw in §4.2.4, even some of the tracking uses of tēnei in this text have an affective force.

5.2 Personal Narrative 2 (PN2)

This narrative, by contrast, is a dramatic, well-known story with a coherent narrative shape. It has been told many times, so that there is an element of pre-planning, and the speaker is fluent. While fully aware of the drama of his tale he does not seem so emotionally involved as the other speaker, perhaps because the events took place over 25 years earlier. He does however have an agenda, which is to show how his actions were
validated by subsequent events, and to this end he uses repetition, parallel constructions, and demonstratives, though not as freely as in PN1. Discourse deixis accounts for half of all demonstratives and includes emphatic or focus constructions, as in (17) and (23) above. In another favourite construction of this narrator, first mention of a referent is followed by a co-referential noun phrase in an equational clause with a demonstrative.

(43) *Fakatau taku apa ika i te lua o aho. Hēai buy 1s.POSS.AL can fish LOC SPEC two POSSINAL day NEG he mea e kiki ai. Na maua lā taku popo NSPEC thing TA eat.with ANAPH TA get INTSF 1s.POSS.AL coconut i taku havalivaliga. Ia. Ko te popo lā ia LOC 1s.POSS.AL walk:RDP.NR D.PTCL PR SPEC coconut INTSF DEM ma te apa ika na ola ai au. and SPEC can fish TA live ANAPH 1s ‘I bought a can of fish on the second day. Not a thing to eat with it. But I found a coconut on my walking. It was that coconut and the can of fish that I lived on.’ (PN2:11)

In this example the distinction between tracking and discourse deixis is of less interest than the way in which the demonstrative is being used for dramatic purposes.

5.3 Folktales (TT and TTT)

The Folktales (TT and TTT) are collected in Huntsman (1977 and 1980). The introductions to these volumes give a detailed account of the style of the tales and the skills of the narrators. I examined five folktales, told by four different narrators, amounting to over 11,000 words, and draw on some others in my examples. On the whole, demonstratives are sparsely used. One tale has none, most have just a few, averaging one demonstrative in 200–300 words. Only The tale of Tautilifaga contains a level of demonstratives (1/88) which is similar to that of the other genres in the study. As might be expected in a story, tracking uses account for the vast majority of tokens and there is by far the lowest incidence in the corpus of discourse deixis. The Tokelau folk tales are told with great verve, with a skilful use of repetition, song and reported speech. Nevertheless, the language of the tales is characterised by a notable lack of affect; indeed, horrendous acts of mutilation or murder are described in the most offhand manner, and emotions are seldom referred to.

5.4 Sermon (S)

The Sermon (S) is a highly repetitive text on the subject of the prodigal son. The demonstratives are distributed over 63% discourse deixis and 37% reference tracking. The latter may be rather a high figure for this genre, and is due to the retelling of the story of the prodigal son.
5.5 Formal Speech (UN)

The Formal Speech (UN), welcoming members of the United Nations Committee of 24, is a 526-word text containing nine demonstratives: six tokens of tēnei in situational deixis, referring to the time and place of the meeting (see (25) above), and three examples of discourse deixis, all entirely unremarkable. This text is too short for any reliable conclusions to be drawn.

5.6 Village Council Meeting (VC)

The Village Council Meeting (VC) contains a very high number of demonstratives (1/38) but in a distribution which reflects the dialogic nature of this text. Although there is plenty of discourse deixis, the high overall rate is accounted for by the very high level of situational deixis, expressed mostly by proximal forms referring to time, place, participants and village institutions: te vāihao tēnei (SPEC week DEM) ‘this week’, ki komiti tēnei (to committee DEM) ‘to this committee’. In addition, forms referring to the actual topics or sections of the discussion, which have been counted as text or discourse deixis as appropriate, are also in a sense situational. Himmelmann makes a similar point concerning self-reference to a story or book (1996:221). There are many such phrases.

(44) Ko te tāloha tēnei a Foua

PR SPEC tēnei DEM POSS.AL Foua

‘This is Foua’s request’ (VC:145)

5.7 Tokelau fishing texts by Peato Perez (PTF and PTT)

The two fishing texts (PTF and PTT) are didactic in nature and full of passionate persuasiveness. Hooper and Huntsman (1991) refer to the importance of fishing in Tokelauan life, extending far beyond its purely economic value to matters of status, reputation, cultural identity and emotional satisfaction. They record how Perez ‘prepared brief notes …. which he then [orally] recorded carefully and deliberately’, and they describe him as ‘a consummate Tokelau elder — an orator, preacher and moralist’ (1991:253). Considerable thought has been devoted to the organisation of these texts about the preservation of traditional fishing techniques, a matter on which he has strong feelings. A higher concentration of demonstratives is found here than in any of the other texts except the Village Council. Discourse deixis is by far the predominant use, 76% overall, and in some extensive passages over 90%. Many passages showing Perez’s deployment of demonstratives are cited in this paper, including examples (6), (14), (15), (18), (20), (22) and (41).

I have written previously (Hooper 1988:51–71; 1993:212–227; see also Prince 1978) about the rhetorical potential of Tokelauan equational sentences with demonstrative pronoun absolutes and extrapolated relative clauses, as in examples (18) and (22). These are equivalent to English it-clefts. Uniquely in this corpus (and also in his other works about Tokelau history and politics), Perez shows great skill in using cleft constructions for his own purposes. Note that in (45) the old information is presented as focus (ko te taimi tēnā) and the normally presupposed clause contains what is new and important.
(45) Ko te taimi e hī mai ai te atu, ko te taimi tēnā e fufuti ai tau mānava, oi taofi.
   PR SPEC time TA fish DIR ANAPH SPEC skipjack PR SPEC time DEM
   TA haul ANAPH 2s.POSS.AL breath SQ hold
   ‘At the time that you pull in the skipjack, it is at that moment that you draw in your breath, and hold it.’ (PTF:81)

Example (46) contains a presuppositional clause which expresses an assumption on the part of the speaker which he is presenting as known fact, although it may be highly debatable: he begs the question of whether everything has indeed ‘fallen apart’.

(46) Ko te ala lava ia kua holoina ai ia mea uma lele, nā mea tau ki tautai ma mea tau
   PR SPEC reason INTSF DEM TA collapse ANAPH ABS thing
   all INTSF SPEC.PL thing concerning to fisher and thing concerning
   ki tuākau.
   to ocean
   ‘It is for this reason that absolutely everything has disintegrated, customs relating to master fishermen, and customs relating to the ocean.’ (PTT:1)

5.8 Noosing Wahoo text (NW)

Noosing Wahoo (NW) is another text concerned with fishing skills, a record of an interview session between a researcher and several Tokelau elders. It makes an interesting comparison with the Perez texts. First, the discussion was unplanned. Second, although full of prescriptions and technicalities, it has none of the emotional force of Perez’s urgent calls for excellence. The functions of demonstratives are distributed more evenly than in any other text: there are almost as many (hypothetical) tracking uses as discourse uses, and some situational uses referring to participants and local people.

6 Conclusion

Certain associations of demonstrative functions with individual texts are not at all surprising: situational functions in the formal speech and the village council meeting, or tracking functions in the folktale. What is less expected is the high proportion of discourse deixis in all texts except the folktale, and the extremely high proportion in the most emotionally charged texts, PN1, PTF and PTT. Although PN2 has comparatively few demonstratives, over half of them are discourse deixis, but used in a more artfully rhetorical way than in PN1. As I pointed out, this narrator had practice in polishing his tale. The most pervasive and skilful deployment of demonstratives is found in the Perez texts, resulting both from the force of feeling displayed and the rhetorical skills honed by a lifetime of teaching and preaching.

References


1 Introduction

Oceanic languages typically contain a type of construction, or more accurately, a type of lexical item, known as object incorporation. In object incorporation, a noun that would otherwise function as the head of a direct-object noun phrase forms a lexical item together with the relevant verb, a kind of compound verb. Example (1) from Toqabaqita shows the noun qai ‘(fire)wood’ as the head of a direct-object noun phrase:

(1) Ni nau na kwai ngali-a mai qai lakoo ki.  
PROFORE 1s FOC 1s.FUT carry-3.OBJ VENT wood that PL 
‘It will be me who will bring those pieces of (fire)wood.’

Note that the direct object qai lakoo ki ‘those pieces of firewood’ is separated from the verb ngali ‘carry’ by the ventive directional mai, and that it contains a demonstrative and the plural marker. The object is definite and is indexed on the verb ‘carry’ by means of the suffix -a.

In (2), on the other hand, the noun qai ‘wood’ occurs immediately after the verb ‘carry’, before the ventive directional. The noun qai cannot be modified by a demonstrative, the plural marker, or any other kind of modifier found in noun phrases. Finally, the verb ‘carry’ does not have the object-indexing suffix -a. There is no direct object in (2).

It gives me great pleasure to offer this paper to Andrew Pawley. Over the years, Andy has been a teacher, a colleague, and always a good friend. His studies of various aspects of Oceanic grammar have been a great source of inspiration for me. And it was Andy, together with George Grace, who suggested that I do field work on the Manam language, as part of my PhD studies at the University of Hawaii. This paper was written while I was an Honorary Visiting Fellow at the Research Centre for Linguistic Typology at La Trobe University, February – July 2006. I am grateful to R.M.W. Dixon and Alexandra Aikhenvald for making it possible for me to enjoy the intellectual atmosphere at the Centre. A version of the paper was delivered at the University of Cologne in December 2006. I am grateful to a number of those present for comments that I took into account when preparing the final version. My thanks also to the anonymous referee for comments on the original version from which the paper profited in a number of ways.

In agreement with Mithun (1984) and Rosen (1989), I consider such verb-noun combinations in the Oceanic languages to be the result of word-formation processes rather than the result of a syntactic process of incorporation, as argued by, for example, Baker (1988).
In (2) the verb *ngali* ‘carry’ and the noun *qai* ‘wood’ form a compound, an intransitive verb. The evidence that the compound verb is intransitive can be mentioned only briefly here. First, when such verbs are nominalised, the possessor of such a nominalisation corresponds to the subject of the corresponding verb, which is the case with simple intransitive verbs. On the other hand, in the case of nominalisations of transitive verbs, with qualifications that need not concern us here, the possessor corresponds to the direct object of the corresponding verb. Secondly, such compound verbs participate in verb-verb compounds (‘verb serialisation’) in the way simple intransitive verbs do.

Object incorporation that results in intransitive verbs is the dominant pattern in Oceanic. There are some languages that have object incorporation that results in transitive verb-noun compounds. Toqabaqita is one such language. In (3) the transitive verb *qui* ‘strike’ has as its direct object *si onga nia* ‘his tattoo’:

(3) \[\text{Wane kere qui-a si onga nia qana iqa.}\]
\[\text{man 3p.NFUT strike-3.OBJ CLF tattoo 3s PREP fish}\]
\[\text{‘The man had a tattoo of a fish done.’ (lit. ‘The man, they struck his tattoo of fish.’)}\]

In (4) below the verb ‘strike’ and the noun ‘tattoo’ form a compound. Instead of *qui*, the verb has the form *quu*, which is the intransitive counterpart of *qui* and which also serves as a ‘combining’ form of *qui* (used, for example, in serialising verb-verb compounds). The noun ‘tattoo’ does not have the classifier *si*, and there is no possessor phrase present. The compound verb has a direct object that refers to the part of the body that is tattooed. That direct object is indexed on the compound verb by means of the object suffix -*na*.

(4) \[\text{Nau ku thathami-a qoki quu-onga-na fasi qaba-ku.}\]
\[\text{1s 1s.NFUT want-3.OBJ 2s.FUT strike-tattoo-3.OBJ PREC arm-1s.PERS}\]
\[\text{‘I want you to tattoo my arm.’}\]

In both types of object-incorporating compounds the order of the constituents is verb-noun, which corresponds to the order of the verb and its direct object in a verb phrase.

Finally, there is a third type of compound verb, one that involves a verb and an incorporated noun, where the incorporated noun corresponds to the head of the subject phrase of the base verb. In (5) the verb is *dora* ‘forget’ and its subject is *lio-ku* ‘my mind’. The ‘possessor’ of the mind is also encoded by means of the first person singular independent pronoun *nau*. However, the pronoun does not occupy the possessor position in a possessive noun phrase; rather, it occupies the topic position. (In possessive noun phrases the possessor phrase follows the possessum phrase; see *thata-na wane qeri* (name-3.PERS man that] ‘the man’s name’ in (5).) I will return to the topic status of possessors in §4.

---

3 There are two basic classes of transitive verbs in the language. Verbs of one class index their direct objects by means of the suffix -*a*, while verbs of the other class index their direct objects by means of -*na*. Compare *qui-a* in (3) and *quu-onga-na* in (4). All object-incorporating verbs that are transitive are members of the latter class.
Subject incorporation in Toqabaqita and other Oceanic languages

In (5) the subject marker is third person singular non-future, e. It indexes the subject lio-ku ‘my mind’. In (6) below, the noun ‘mind’ and the verb ‘forget’ form a compound. The noun cannot carry a personal suffix indexing the possessor (cf. -ku 1s.PERS in (5)), and the subject marker is first person singular non-future, ku, indexing the subject, ‘I’.

(6) Nau ku lio-dora qana thata-na wane qeri.
    1s 1s.NFUT mind-forget PREP name-3.PERS man that
    ‘I have forgotten the man’s name.’ (lit. ‘I have mind-forgotten the man’s name.’)

Lio-dora in (6) is a subject-incorporating compound. Example (8) further below also contains a subject-incorporating compound verb, while (7) contains the corresponding syntactic, analytic construction with the noun maa ‘eye’ in the subject phrase. In (7) the subject marker is third person singular, e, while in (8) the subject marker is second person singular, qo.

(7) Maa-mu e geqo.
    eye-2s.PERS 3s.NFUT be.’blind’
    ‘Are you blind?’

This could be said jokingly to a person who has failed to see something that is in full view, lit. ‘Your eyes are ‘blind’.’ (Geqo signifies a condition when a person is partially or fully blind(ed) because of an eye sickness.)

(8) Qo maa-geqo.
    2s.NFUT eye-be.’blind’
    ‘Are you blind?’ (lit. ‘You are eye-blind.’). (Used in the same way as (7).)

In subject-incorporating compounds the order of the constituents is noun-verb, which corresponds to the order of the subject and the verb.

It is subject incorporation that is the focus of this study, but some comparison between subject incorporation and object incorporation will be made (in §4). Section 2 discusses in more detail subject incorporation in Toqabaqita, and §3 looks at subject incorporation in some other Oceanic languages. Section 4 considers the motivation for subject incorporation. Some more general comments are made in §5.

2 More on subject incorporation in Toqabaqita

A number of researchers have pointed out that subject incorporation is restricted in various ways. First, it normally applies only to S noun phrases, not to A noun phrases; second, it is typically restricted to verbs whose S arguments are non-agentive, normally patient or theme; and third, it typically takes place with nouns that designate body parts (see, for example, Mithun 1984; Baker 1988; Rosen 1989). Verbs that undergo subject incorporation are sometimes said to belong in a class of unaccusative verbs, and in some syntactic theories the surface S argument is assumed underlingly to occupy the direct-object position (e.g. Baker 1988). The term ‘body part’ will be interpreted somewhat
loosely here: the category may include the noun for ‘body’ and nouns for more abstract concepts such as the mind.

In Toqabaqita, the range of nouns that can participate in subject incorporation is quite restricted: ‘body’ (the most common one), ‘mind’ (two nouns), ‘belly (as the seat of certain emotions)’, ‘breath’, ‘eye’, ‘head’, ‘mouth’, and a few others in more idiosyncratic combinations. The list in (9) below gives some indication of the kinds of subject-incorporating compounds found in the language. It is not meant to be exhaustive. The list is divided into sections according to the incorporating noun. All of the compounds given in (9) have synonymous counterparts where the relevant noun is not incorporated and occurs in the subject position with that verb. In many cases the base verb occurs only with that noun in the subject position, and so the meanings assigned to such verbs are, somewhat arbitrarily, assumed to be the same as those of the corresponding compounds. Such verbs are marked with 

(9) a. seqe ‘body’

- seqe-daadaola ‘be/feel lazy’, ‘not feel like doing s.th.; daadaola ‘of one’s body or a body part: be tired, stiff, numb (after exertion or long inactivity)’
- seqe-lufa ‘be shy, embarrassed, ashamed’; lufa#
- seqe-noni ‘be shy, embarrassed, ashamed’, same as seqe-lufa; noni#
- seqe-raaqa ‘not have inhibitions, be uninhibited about doing s.th.’;
  ‘not be shy, embarrassed, afraid to do s.th.’; raaqa#

b. lio and manata both ‘mind, thought’

- lio-dila ‘be, feel very sad, dejected, heartbroken’; dila ‘slip, slide’
- lio-dora ‘forget’; dora ‘forget’, ‘not know’
- manata-dora ‘forget’, same as lio-dora
- manata-akele ‘repent’, ‘regret having done s.th. wrong’; akele ‘turn (around, over)’
- manata-naqirua ‘be undecided, be of two minds’; naqirua# (cf. rua ‘two’)
- manata-ruarua ‘be undecided, be of two minds’, same as manata-naqirua; ruarua# (cf. rua ‘two’)

c. rake ‘belly’

- rake-angongo ‘be very angry’; angongo#
- rake-boko ‘be constipated’; boko ‘be shut, blocked’
- rake-qiri ‘be angry’; qiri#

d. gwau ‘head’

- gwau-boko ‘be dumb, not smart, not intelligent’; boko ‘be shut, blocked’
- gwau-sadi ‘be dumb, not smart, not intelligent’, same as gwau-boko; sadi ‘be hard (not soft)’

e. maa ‘eye’

- maa-boko ‘be blind, blinded, unable to see, because of permanent blindness or a temporary eye problem’; boko ‘be shut, blocked’
maa-geqo ‘be ‘blind’ (in a jocular sense)’ (see (8) above); geqo ‘of eyes:
be partially or fully blind(ed) because of a sickness’
maa-qasia ‘avert gaze, avoid looking at s.o., by looking down’; qasia ‘fall down’
f. mango ‘breath’
    mango-thaqethaqe ‘pant; be out of breath’; thaqethaqe
    mango-una ‘be short of breath (e.g. after exertion), ‘suffer from shortness
    of breath (medical condition)’; una

g. foko ‘mouth’
    foko-qato ‘be mute; not have the faculty of speaking (such as congenitally)’;
    qato
Verbs that involve subject incorporation are fully-fledged verbs and may themselves serve
as bases for derivations. For example, besides seqe-nonì ‘be shy, embarrassed, ashamed’ in
(9a) above there is a deverbal noun seqe-nonì-a ‘shyness, embarrassment, shame’. Similarly, besides rake-qiri ‘be angry’ in (9c) there is a deverbal noun rake-qiri-a ‘anger’,
and the form gwau-boko in (9d) functions not only as a verb, ‘be dumb, not smart, not
intelligent’, but also as a noun, ‘dumb, unintelligent person; blockhead’.
There are also some subject-incorporating compounds that are not synonymous with
their syntactic, analytic counterparts, without incorporation. A few examples are given in
(10)–(12), where the (a) forms are subject-incorporating compounds, and the forms in (b)
are the corresponding analytic constructions. The meanings of the constituents are given in
(c).

(10) a. seqe-thaathala ‘be/feel happy’
    b. seqe … thaathala ‘feel good physically’
    c. seqe ‘body’, thaathala ‘be light in weight’

(11) a. seqe-leqa ‘be new’, ‘be clean’
    b. seqe … leqa ‘of a body: be good, nice’
    c. seqe ‘body’, leqa ‘be good, nice’

(12) a. ifu-kulu ‘have long, unkempt hair (one of the traditional
    manifestations of being in mourning)’
    b. ifu … kulu ‘of (long) hair: hang down (no implication of the
    person being in mourning)’
    c. ifu ‘hair’, kulu ‘hang down’

4 As in:
    (i) Kaleko seqe-n=leqa.
    clothes body-3s.PERS=3s.NFUT be.nice
    ‘The clothes are nice.’ (lit. ‘The clothes, their body is nice.’)

5 As in:
    (ii) Kini kwao, ifu-na e kulu.
    woman be.white hair-3s.PERS 3s.NFUT hang.down
    ‘The white (Caucasian) woman has (long) hair hanging down.’
    (lit. ‘The white woman, her hair hangs down.’)
There are also a few noun-verb compounds where the noun corresponds to the subject of the verb, but they do not have syntactic counterparts of the type discussed above. They are not directly related to the topic at hand and their existence is merely noted. One such compound is \textit{oga-fii} ‘of a coconut: be somewhat rancid, past the best stage for drinking’; cf. \textit{oga} ‘intestine’ and \textit{fii} ‘hurt, be painful’. This verb takes as its subject noun phrases that refer to coconuts:

(13) \textit{Niu neqe e oga-fii naqa, ....}\newline\textit{coconut this 3s.NFUT intestine-hurt PRF}\newline‘This coconut is somewhat rancid, not good for drinking, ....’

The incorporated noun ‘intestine’ is not in a body-part relation to ‘coconut’.

Another compound is, in fact, a noun, not a verb: \textit{kwara-fita} ‘the end of the urinary tract where the urine leaves the body’; cf. \textit{kwara} ‘urine’ and \textit{fita} ‘run’.

There is at least one subject-incorporating compound that has a synonymous analytic counterpart and that is exceptional with respect to the pattern discussed above: \textit{akalo-taqe} (‘ancestral.spirit-ascend’), an intransitive verb whose meaning is ‘of a person: have a visitation or communication from an ancestral spirit’:

(14) \textit{Wane e akalo-taqe.}\newline\textit{man 3s.NFUT ancestral.spirit-ascend}\newline‘The man had a visitation/communication from an ancestral spirit.’

The verb \textit{taqe} has a number of senses, the primary ones being ‘move, go up; ascend (moving on the ground or rising into the air)’. The analytic counterpart of \textit{akalo-taqe} contains not the intransitive verb \textit{taqe} but its transitive counterpart \textit{taqe-li}, whose senses include ‘climb (to the top of)’, and, importantly for the present discussion, ‘of an ancestral spirit: visit someone and communicate (s.th.) to him/her’. The analytic counterpart of (14), with the subject-incorporating compound \textit{akalo-taqe}, is shown in (15). (The basic constituent order in transitive clauses is AVO.)

(15) \textit{Akalo e taqe-li-a wane.}\newline\textit{ancestral.spirit 3s.NFUT ascend-TR-3.OBJ man}\newline‘An ancestral spirit visited and communicated (s.th.) to the man.’

In (16) the noun phrase that encodes the person visited by a spirit has been topicalised and fronted:

(16) \textit{Tha weleqi, akalo e taqe-li-a.}\newline\textit{PERSMKR guy ancestral.spirit 3s.NFUT ascend-TR-3.OBJ}\newline‘The guy was visited by an ancestral spirit.’ (lit. ‘The guy, an ancestral spirit ascended him.’)

The subject-incorporating verb \textit{akalo-taqe} is exceptional on two counts. First, the incorporated noun corresponds to the A argument of a transitive verb, and secondly, the incorporated noun does not designate a body part (in the sense defined above) with respect to the type of entity encoded in the subject position (‘man’ in (15) and (16)). As discussed in §3 example (27), there is another language where, exceptionally, the incorporated noun corresponds to the A argument of a transitive verb, but there the subject-incorporating compound itself is transitive.
This concludes the discussion of the formal and the basic semantic aspects of subject-incorporation in Toqabaqita. In the next section we will take a brief look at subject incorporation in a few other Oceanic languages.

3 Subject incorporation in other Oceanic languages

By and large, judging by the paucity of explicit discussions of subject incorporation in the descriptions of Oceanic languages, the phenomenon either does not exist at all in many of them, or is only marginal and so may have escaped notice. One language where it is found and for which it has been discussed in detail, is Saliba. Margetts (1999) discusses Saliba subject incorporation under the rubric of ‘external possession’:

In Saliba EP [external possession; FL] constructions, an entity which is classified in the language as inalienably possessed is incorporated into the verb and the semantic possessor appears as the subject of the construction. [Footnote omitted.] The corresponding analytic constructions show an intransitive verb with a possessed noun as its subject. That is, the possessor in the EP constructions holds the same syntactic relation to the verb as the possessed noun in the corresponding analytic construction. (Margetts 1999:233, emphasis added.)

While only nouns that express inalienably possessed concepts can participate in subject incorporation in Saliba, not all of them can: ‘Only person part terms (i.e. body part terms and more abstract concepts like kamna ‘feeling’ or nuwa ‘mind’) occur as incorporated nouns in Saliba EP constructions.’ (Margetts 1999:234).

Examples (17) and (18) show, respectively, an analytic construction with the possessum expressed in the subject position and the corresponding compound with the possessum noun incorporated:

(17)  Boga-gu ye-kamkamna.
      belly-1s.POSS 3s-hurt
   ‘My belly hurts.’

(18)  Ya-boga-kamkamna.
      1s-belly-hurt
   ‘I have a stomachache.’ (Margetts 1999:236)

Some other subject-incorporating compounds in Saliba are: gado-bigga (‘throat-soft’) ‘not be thirsty’, beya-kolakola (‘ear-deaf’) ‘be deaf’, kamna-yababa (‘feeling-bad’) ‘feel bad’, nuwa-polohe (‘mind-heavy’) ‘be undecided’. According to Margetts (1999:236), ‘[t]he stem nuwa ‘mind’ is one of the most frequent nouns to be incorporated’. In Saliba, subject incorporation is lexically restricted. A given verb may form a compound with certain body-part nouns but not with others. For example, the following combinations exist: ‘belly-hurt’, ‘head-hurt’, ‘back-hurt’, ‘eye-hurt’ and ‘ear-hurt’, but for example, ‘foot-hurt’ and ‘mouth/tooth-hurt’ do not, even though there are corresponding analytic constructions.

In many cases a subject-incorporating compound and the corresponding analytic construction are (broadly) synonymous, but there are also cases where the meaning of the compound is different from that of the corresponding analytic construction:
Subject incorporation is also found in Tawala. Ezard (1997:278) calls such compounds ‘human-propensity verbs’, because ‘[t]hey match closely the ‘human propensity’ adjectival class of Dixon (1982:19), but in Tawala they are clearly verbal ….’. These verbs consist of a body-part root (including ‘mind’) and a verb root, in that order. The verbs that enter into human-propensity compounds are mostly stative, but some active verbs do occur (such as -guluwi ‘bury s.th.’). Not all subject-incorporating compounds have analytic counterparts, but when they do, the difference, according to Ezard, often has to do with the degree of control the referent of the subject or of the possessor expression has over the relevant state of affairs. The subject-incorporating form is used when the person has some control over the state of affairs, while the analytic construction is used when the person has no control, as in (21) and (22), respectively:

(21)  A-upu-dodola.
     1s-back-stiff
     ‘I am lazy.’ (some control)

(22)  Upu-we i-dodola.
     back-1s  3s-stiff
     ‘I am lazy.’ (no control) (Ezard 1997:101)

Since subject-incorporating forms are used when the relevant person has some control over the state of affairs, such verbs can occur in imperative sentences, while the corresponding analytic constructions cannot:

(23)  Ipa o-na-nugo-kadidili.
     IRR  2p-POT-mind-hard
     ‘Be courageous.’

(24)  *Ipa nugo-m i-na-kadidili.
     IRR  mind-2s  3s-POT-hard
     (‘Let your mind harden.’) (Ezard 1997:101)

According to Ezard (1997:100), the use of human-propensity verbs is ‘[a] highly productive area of Tawala grammar’, although it is not clear whether the productivity relates to the frequency of use of those verbs or to their formation, or to both. Many such compounds are ‘idiomatic’ (Ezard 1997:279), which suggests that they are not synonymous with their analytic counterparts. For example:

(25)  I-taniga-pota.
     3s-ear-block
     ‘He is disobedient.’
There is at least one case where the incorporated noun corresponds to the A argument of a transitive verb. The compound too is transitive.

(27) \[ A\text{-}nugo\text{-}guluw'\text{-}i. \]
\[ 1\text{s-mind-bury-3s} \]
‘I forgot it.’

Compare:

(28) \[ Nugo\text{-}u \text{ i-guluw'\text{-}i.} \]
\[ \text{mind-1s 3s-bury-3s} \]
‘I forgot it.’ (Ezard 1997:101)

The Tuvaluan language has what Besnier (2000:606) calls ‘verbal noun-adjective compounds’, which consist of ‘a noun which denotes a body part (e.g. gutu ‘mouth (area)’, …), a centre of cognition or of emotion (e.g. loto ‘heart’, …), or the product of a mental or body function (e.g. leo ‘voice’, …), modified by an adjective or verb denoting a quality or activity associated with the body part or function denoted by the first element’. Although Besnier calls the compounds ‘noun-adjective’ compounds, he says elsewhere in the grammar (2000:260) that there are no structural criteria that separate the categories of verbal adjectives and verbs. The compounds ‘function like regular verbal adjectives’ (2000:607). The Tuvaluan compounds resemble the subject-incorporating compounds in the languages discussed above, although Besnier does not specifically discuss the existence of analytic counterparts of such compounds. Three examples of verbal noun-adjective compounds are given in (29):

(29) a. \[ mata \ faanoanoa \]
\[ \text{eye sad} \]
‘sad looking’

b. \[ loto \ alofi \]
\[ \text{heart feel.empathy} \]
‘empathetic’

c. \[ gutu \ ppelo \]
\[ \text{mouth lie} \]
‘prone to lying’ (Besnier 2000:606–607)

Besnier describes the meanings of verbal noun-adjective compounds thus:

These compounds tend to denote characterising traits. For example, someone is said to *ppelo ‘lie’ if he or she is caught lying once, but someone who is *gutu *ppelo [see (33c) above; FL] is a habitual liar. (Besnier 2000:607)

As in all the other languages discussed above, the order of the constituents in the compounds is noun-verb. Tuvaluan is verb initial, but it does allow other orderings of the main clause constituents.

Noun-verb compounds that correspond to subject-incorporating compounds in other languages are also found in Samoan, although Mosel and Hovdhaugen (1992) do not
establish the category ‘subject’ for the language. Mosel and Hovdhaugen call such compounds ‘isu mamafa compounds’ (isu ‘nose’, mamafa ‘heavy’):

*Isu mamafa* compounds are noun-verb compounds in which the verb modifies the noun. The noun denotes a body part or some other part of a whole, and the entire compound means ‘to have a (body)part of the quality expressed by the modifying verb’. Syntactically, *isu mamafa* compounds behave like non-ergative verbs. (Mosel and Hovdhaugen 1992:300)

*Isu mamafa* compounds can function predicatively or attributively; see (30) and (31) for the former and (32) for the latter:

(30) ‘Ua ‘ou isu mamafa.
PRF 1s nose heavy
‘I have a cold.’ (lit. ‘I have a heavy nose.’) (Mosel and Hovdhaugen 1992:300)

(31) Na iloa e Maatusi ua loto vaivai Iopu.
PST know ERG Maatusi PRF heart weak Iopu
‘Maatusi recognised that Iopu was scared.’ (lit. ‘… that Iopu had a weak heart.’) (Mosel and Hovdhaugen 1992:335)

(32) se puaa nifo tele
ART.NSPEC.PL pig tooth big
‘a pig with big teeth’ (Mosel and Hovdhaugen 1992:300)

4 The motivation for subject incorporation

We can now turn our attention to the function of subject incorporation. The discussion will focus on Toqabaqita, because this is the language on which I have the most detailed information. To anticipate, the function of subject incorporation is to topicalise the (notional) possessor and so to make that participant (more) salient. By topic is here meant the (type of) entity that the rest of the sentence is a statement, a question, etc. about.

Before discussing the function of subject incorporation, it is useful to take a quick look at object incorporation. Object incorporation of the typical Oceanic type is usually said to take place when the direct object would be non-referential, generic. This is the case in (2), presented again below as (33):

(33) Ni nau na kwai ngali qai mai.
PROFORE 1s FOC 1s.FUT carry wood VENT
‘It will be me who will bring (fire)wood.’

Object incorporation is used when a sentence encodes a general, habitual type of event directed at a certain type of entity rather than a specific event directed at a specific entity:

(34) Manga na=i dini maka nau kai baa-basi futa
time REL=LOC distant.past father 1s 3s.IPF RDP-shoot possum
qana basi naqi.
PREP bow this
‘In the old times, my father used to shoot possums with this bow.’

---

6 The literal glosses in (30) and (31) are Mosel and Hovdhaugen’s.
In (34) *baa-basi futa* ‘shoot possums’ is an object-incorporating compound: the verb ‘shoot’ does not carry an object-indexing suffix, and nothing could intervene between the base verb and the noun.

However, while non-referentiality of what otherwise would be the direct object is sufficient to trigger object incorporation, it is not a necessary condition. The question in (35) can be addressed to people who are, at that moment, picking canarium nuts (knocked off a tree to the ground):

(35)  
\[ \text{Muki gwee ngali?} \]
\[ \text{2p.IPF pick canarium.nut} \]
\[ ‘\text{Are you picking canarium nuts?}’ \]

Here there are specific entities, specific canarium nuts involved. Nevertheless, object incorporation has taken place. What is important is that the exact identity of the canarium nuts is not relevant to the discourse at hand. The nuts are not discourse salient. The question is about what the people are doing, not about the identity of the nuts being picked. In fact, interrogatives of this type are routinely used as a kind of conversation opener, rather than as requests for information.

Even a noun designating a body part/body parts can be incorporated, even though the identity of the body part(s) is obvious. In (36) there is no object incorporation. The sentence is transitive, with the noun phrase ‘my teeth’ as its direct object:

(36)  
\[ \text{Nau kwai dada-a fasi lifo-ku.} \]
\[ \text{1s 1s.FUT brush-3.OBJ PREC tooth-1s.PERS} \]
\[ ‘\text{I am going to brush my teeth now.’} \]

However, it is possible to incorporate the noun ‘tooth’, even though there is no doubt about the identity of the teeth that will be brushed. Note the absence of the object suffix in (37) compared to its presence in (36), and the positioning of the precedentive particle *fasi* after *dada lifo*, rather than between the two, as in (36).

(37)  
\[ \text{Nau kwai dada lifo fasi.} \]
\[ \text{1s 1s.FUT brush tooth PREC} \]
\[ ‘\text{I am going to brush (my) teeth first.’} \]

While, in principle, the speaker might intend to brush some unspecified teeth, the sentence is perfectly fine if the speaker intends to brush his/her own teeth, and would be normally interpreted as such. Here too it is not lack of referentiality, but lack of discourse salience that is the relevant factor.

It is lack of discourse salience, of which nonreferentiality is just a special case, that is the relevant factor in object incorporation.

As discussed in §2, in Toqabaqita, subject incorporation typically takes place with body-part nouns:

(38)  
\[ \text{Wela e rake-boko.} \]
\[ \text{child 3s.NFUT belly-be.blocked} \]
\[ ‘\text{The child is constipated.’} \]

For the corresponding analytic construction see (39) below.

There is no uncertainty in (38) over which/whose belly is blocked. Here too nonreferentiality is not the relevant factor. Even though it is the possessum that is affected
(here the child’s belly), ultimately it is the child himself/herself that is affected. The belly is only the medium through which the child is affected. As Bally (1996) puts it, body parts are part of their possessors’ personal domains, and ‘... each phenomenon, action, state or quality which affects any part whatsoever of the personal domain, automatically affects the whole person.’ (Bally (1996:33).

In Toqabaqita, subjects are default topics. Unless another sentence element has been topicalised, it is the subject that encodes the topic. Subject incorporation of the type discussed here is a kind of external possession construction (Payne and Barshi 1999a; see also various contributions in Payne and Barshi 1999b; and Margetts 1999 for Saliba, as discussed in §3 above). In this type of external possession, it is the notional possessor that is expressed as the subject of the sentence and hence as the (default) topic. Sentences with subject incorporation are primarily statements, questions, etc. about the notional possessors, not about their body parts.

Subject incorporation is not the only strategy by which topicalisation of the possessor can be achieved. There is another strategy that involves topicalising the possessor, but without making it the subject of the sentence and without incorporation of the possessum. In (38) above, the possessum ‘belly’ has been incorporated and the external possessor ‘child’ is in the subject position. On the other hand, in (39) below, which is broadly synonymous with (38), the noun ‘belly’ still occurs in subject position, while the expression of the possessor, ‘the child’, has been fronted into topic position. (The normal position for the possessor phrase is to follow the possessum phrase; see (5) in §1 and the paragraph before it.) The topic is indexed on the subject noun by means of the third person singular personal suffix -n. (The full form of the suffix is -na, but because of the cliticisation of the third person singular non-future subject marker, its form is -n in (39).)

(39) Wela rake-n=e boko.
child belly-3s.↑be.blocked
‘The child is constipated.’ (lit. ‘The child, his/her belly is blocked.’)

Even though the noun ‘belly’ is in the subject position, the sentence is not a statement about the child’s belly, but about the child.

The pair of examples in (40) and (41) also show the two alternative strategies:

(40) Kini e foko-qato.
woman mouth-be.mute
‘The woman is mute.’

(41) Kini foko-n=e qato.
woman mouth-3s.↑be.mute
‘The woman is mute.’ (lit. ‘The woman, her mouth is mute.’)

The subject-incorporating strategy and the topicalising strategy have the same effect: topicalisation of the possessor. Note that this kind of topicalisation is not necessarily triggered by considerations of discourse continuity, although it may be. That is, it is not triggered just by the fact that the possessor is the current discourse topic. Constructions with subject incorporation of the type shown in (38) and (40) and topicalisation of the possessor of the type shown in (39) and (41) are, in fact, the preferred types of construction without any preceding linguistic context, and are normally offered as translations of sentences in elicitation; for example, ‘How would you say in Toqabaqita “The child is
constipated.”?’. This is not to say that sentences without subject incorporation and without possessor topicalisation do not ever occur, as (44) shows. The sentences in (42)–(44) have the same meaning and could be used jokingly in the same way to a person who has failed to see something that is in full view:

(42) Qo maa-geqo.
2s.NFUT eye-be.’blind’
‘Are you blind?’ (lit. ‘You are eye-blind.’) (This is a repetition of (8).)

(43) Qoe maa-mu e geqo.
2s eye-2s.PERS 3s.NFUT be.’blind’
‘Are you blind?’ (lit. ‘You, your eyes are blind.’)

(44) Maa-mu e geqo.
eye-2s.PERS 3s.NFUT be.’blind’
‘Are you blind?’ (lit. ‘Your eyes are blind.’) (This is a repetition of (7).)

However, constructions such as (44), with the possessum noun as the head of the subject phrase, foreground the possessor entity, when in fact it is usually the possessor that is the relevant entity.

Subject incorporation and possessor topicalisation of the kind illustrated here can be thought of as compensating strategies, that is, strategies that ‘… compensat[e] for unfortunate side effects of other categories in the grammar’ (Lucy 1992:59). In Toqabaqita, attributive possessive noun phrases occur in various noun phrase positions, including subject position. Subject position is the default topic position. In some cases a subject possessive phrase does appropriately foreground the possessum entity, as in (45) with a kinship term as its head, and in (46) with a body-part term as its head:

(45) Thaina-maliqa qe mataqi.
mother-1pe.PERS 3s.NFUT be.sick
‘Our mother is sick.’

(46) ... ma mea-na ka thaka mai, ka tarutaru
and tongue-3s.PERS 3s.SEQ come.out VENT 3s.SEQ hang
ba-na qı ruuruqu-na.
LIM-3s.PERS LOC chest-3s.PERS
‘... and her tongue came out [of her mouth] and just hung on her chest.’

The anonymous referee has suggested that subject incorporation has another function, that of ‘lexicalisation’, by which he/she appears to mean formation of compact lexical items that in some cases have idiomatic meanings, different from those of their analytic counterparts. While it is true that subject incorporation results in compact lexical items, the same is true of object incorporation (even though in one type of object incorporation the verb and the noun are, or may be, written with spaces between them). However, this view of subject incorporation would leave unexplained the fact that it is severely restricted: in Toqabaqita, with the one exception shown in (14), subject incorporation is not possible when the relation between the possessum and the possessor is not of the part-whole type, and it is not possible when the possessum is a kinship term (see discussion further below in the main text). In both types of case the possessum has a relatively high degree of salience, a relatively high degree of conceptual independence from the possessor, and so is not incorporated.
Subject incorporation is typically found with terms that are inherently relational, normally body parts, as in the examples given in the preceding sections. Kinship terms, which too are inherently relational, do not participate in subject incorporation; for example, it is not possible to say *‘I am mother-sick’, and mean ‘My mother is sick’. Kinship terms normally designate humans, and the referents of possessive phrases headed by kinship terms are (normally) salient, as salient as the referents of their possessor modifiers. On the other hand, when a part of a person’s body is affected or is in a certain condition, it is the person himself/herself who is normally seen as affected, as being in a certain condition, and it is the person who is given more salience, either through subject incorporation or through topicalisation.

In Tuvaluan, verbal noun-adjective compounds are used to characterise possessors (Besnier 2000; see §3 above). That is, sentences with such compounds are statements, questions, etc. about the persons (the possessors) rather than about their body parts. It is the person, not his/her body part(s), that is salient.

The factor of participant salience can also be seen in Tawala, where, according to Ezard (1997), subject incorporation is used when the possessor participant has a relatively high degree of control over the situation (see §3 above). An entity that has control over a situation is more salient than one that is more of a passive participant in a state of affairs. Participants that have control are foregrounded by being expressed in the subject position rather than in the possessor position.

Langacker (1991:171) has identified several ‘universal cognitive principles’ concerning the degrees of salience of entities that include the following; ‘a whole is more salient than its parts; a physical object is more salient than an abstract entity; and a person has maximal cognitive salience’. The first and the last of these are directly relevant to subject incorporation (as well as to possessor topicalisation of the kind briefly discussed here for Toqabaqita).

5 General discussion

Subject incorporation is found in various Oceanic languages, but it is a lexically highly restricted construction. First, it is typically restricted to body-part terms (broadly defined) as possessums; it does not occur with other inherently relational possessum nouns (kinship terms). Second, while certain body-part nouns and certain verbs may form compounds, not every combination of such nouns and such verbs is possible. For example, in Saliba (Margetts 1999) the combinations ‘belly-hurt’ and ‘eye-hurt’ are possible, but ‘hand-hurt’ and ‘throat-hurt’ are not. However, ‘hand’ can combine with ‘dead’ and ‘throat’ can combine with ‘soft’. With respect to Tamambo, Jauncey (1997:167) says, ‘[subject-predicate compounding] is not a productive process for deriving verbs, and I know of only one lexicalised example’, that being mata-tauhi (‘eye-promise’) ‘choose s.th.’.

Subject incorporation of itself has no effect on verb valency. This is not surprising, because the incorporated noun corresponds to the subject, not the direct object of the base verb.

It is not uncommon for a subject-incorporating compound to have an idiomatic meaning, different from the meaning of the corresponding analytic construction. The noun ‘mind’ in particular appears to participate in such idiomatic compounds.

The use of subject incorporation is functionally/cognitively motivated: it foregrounds, gives salience to, the possessor rather than the possessum. The relatively high degree of
salience of the possessor vis-à-vis the possessum does not derive from the latter not being specific. In fact, the possessum is always specific: it is a part of the possessor. However, sentences with subject-incorporating verbs are primarily statements, etc. about the possessor, not about the possessum.

We then have both functional/cognitive motivation and apparently arbitrary lexical gaps. If a grammatical pattern is motivated, why does it not apply to all the potential candidates? It is, of course, conceivable that there are semantic/pragmatic factors that block full application of a given functionally/cognitively motivated pattern, but unless one can provide an independent characterisation of such factors the argument reduces to vacuity. We may not be able to account, in functional and/or cognitive terms, for all the individual facets of a language but that does not absolve us from searching for explanations.

I have argued that in the case of Toqabaqita, subject incorporation takes place with (certain) terms denoting body parts because when a part of a human body is affected or is in a certain condition, the person himself/herself is affected or is in that condition. In Tawala, according to Ezard (1997), subject incorporation serves to signify a relatively high degree of control over the relevant state of affairs on the part of the possessor. In Tuvaluan, according to Besnier (2000), verbal noun-adjective compounds are used to describe possessors in terms of characterising traits. In all cases, the possessor, the person, has relatively high salience vis-à-vis the possessum, the body part. If this kind of explanation has any validity, it amounts to a cross-linguistic prediction that subject incorporation will not first develop in a language under conditions where the possessor is not highly salient vis-à-vis the relevant possessum. This would be the case with alienable possession; for example, ‘his house’, as in ‘He is big-housed’, that is, ‘He has a big house’. Also, as we saw earlier, subject incorporation does not take place even with kinship terms, where the relation between the possessum and the possessor is of the ‘inalienable’ type. It is, of course, possible, for subject incorporation to become a more general pattern and to spread into those two latter types of domain, but the prediction is that it will not originate in either of them. Obviously, this prediction is open to empirical challenge.

References


The fourth person in Teop

ULRIKE MOSEL

1 Introduction

The Teop language is an Oceanic language spoken in the north-east of the island of Bougainville in Papua New Guinea. It is classified as an Oceanic Meso-Melanesian language of the North-West Solomonic linkage (Ross 1988:213–160, Lynch et al. 2002:101ff.). From a typological point of view Teop shows a number of rare features such as a high number of ditransitive verbs and an extremely complex verb complex with TAM markers, negations, incorporated nouns, adverbs, serial verbs, prepositions, object markers and directional particles (Mosel and Spriggs 1999a:321–349, 1999b:45–56; Reinig 2004:89–105; Mosel forthcoming). Another outstanding grammatical feature is that the coding of arguments is not only determined by the valency of the verb complex and the syntactic functions of arguments, but also by the category of person of the arguments.¹

This paper will show that pronominal and nominal arguments distinguish between two categories of non-speech-act participants which could justifiably be called third and fourth person. To develop our argument, we will first in §2 describe the coding of arguments by cross-referencing morphology and constituent order, and then in §3 turn to the distribution of those two categories of pronouns that refer to non-speech-act participants, and are called the third and fourth person pronouns. In §4 we will show that the category of person also plays a crucial role in the coding of nominal arguments, because the means for marking of nominal arguments with articles parallels the distinction of third and fourth person pronouns. In §5 our findings on the interaction of pragmatics, semantics and syntax in Teop argument structure are summarised.

2 Syntactic functions and coding strategies

Teop has intransitive, transitive ditransitive and ambitransitive verbs. There is no morphological passive, but anticausatives can be derived by the prefix *ta-* from certain inherently causative verbs such as *kaku* ‘break’, *takaku* ‘be broken’. The topicalisation of a

¹ The present work on the Teop language was generously funded by the Volkswagen Foundation and supported by many Teop speakers who helped with recordings, transcriptions, translations and dictionary work. I dedicate this paper to Andrew Pawley who has always showed great interest in our work on the Teop and Samoan languages and helped us to shape our ideas by constructive criticism.
patient or a theme argument of a transitive verb or verb complex can be expressed by word order, see §2.2. Verbs can function as the head of verb complexes which can have a different valency since noun incorporation, serial verbs, the applicative particle or incorporated prepositions can change the valency. In the context of the present investigation, however, valency change is not an issue.

The subject always figures higher on the Actor-Undergoer hierarchy than the object or objects (for Actor-Undergoer hierarchy see Van Valin 2001:32, 2005:58), but with the objects of ditransitive verbs, the argument structure follows a different schema. Formally we can distinguish between the primary and the secondary object. While the primary object is the one that shows exactly the same morphosyntactic features as the sole object of a transitive clause, the secondary object does not (see §2.1 and §3.2). Note that we do not restrict the use of the term ‘primary object’ to objects that are recipients and do not distinguish between primary objects and direct objects (Dryer 1986).

Semantically, three types of ditransitive constructions can be distinguished in Teop.

<table>
<thead>
<tr>
<th>Table 1: Semantic roles of ditransitive verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Type I</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Type II</td>
</tr>
<tr>
<td>Type III</td>
</tr>
</tbody>
</table>

With the first type of ditransitive verbs, the primary object is a recipient as in (1), a location (2) or an addressee (3), and the secondary object a theme or the content of what is said.

(1)  \[
\text{[Me Toko] [paa hee] [bene Sookara] [bona overe].} \\
\text{[and:ART Toko] [TAM give] [OBJ.ART Sokara] [OBJ.ART coconut]} \\
\text{‘And Toko gave Sokara the coconut.’} \quad [\text{Sia. 1.68E}]^2
\]

(2)  \[
\text{[Bana] [e ta kapa] [bono karirava].} \\
\text{[spread] [ART piece metal] [OBJ.ART karirava.leaf]} \\
\text{‘(We) spread the karirava leaf on the piece of sheet metal.’} \quad [\text{Hel. 1.29R}]
\]

(3)  \[
\text{Ei [hivi vai] [anaa] [bono vaabuaku o hivi].} \\
\text{here [ask then] [1s.OBJ] [OBJ.ART second ART question]} \\
\text{‘Here, (she) then asked me a second question.’}
\]

In contrast, with the second type of ditransitive verb the primary object is a patient and the secondary object an instrument (4).

(4)  \[
\text{… me- [paa navu- navuhu ri] [bono nahu] [bona ta naono].} \\
\text{and [TAM RDP- beat 4p] [OBJ.ART.PL pot] [OBJ.ART piece wood]} \\
\text{‘… and (he) beat the pots with a stick.’} \quad [\text{Aro. 2.29.E}]
\]

^2 The abbreviations in square brackets refer to the documents in the corpus of the Teop language project; R stands for the annotated transcription of a recording, E for the edited version of a transcription. All edited versions were done by native speakers (Mosel et al. 2007).
The third type is a causative construction that is derived from a transitive construction (5).

(5) \textit{Vaa- gaga- i bono marasin.}  
\textit{CAUS- drink- 3s OBJ.ART medicine.}  
‘Make him drink the medicine.’ [Sii.Sen. 2.54]

All these kinds of arguments have it in common that they are expressed by bare noun phrases without prepositions and that their semantic role is determined by the valency of the verb complex. In contrast to typical ‘subject languages’, the subject is not an ‘intersection of agent and topic’ (Comrie 1989:66). Rather, the coding of the pragmatic function of topic is independent of that of subject, see §2.2.

Subjects, primary objects and secondary objects are distinguished from each other by an intricate interaction of three coding strategies:

1. the way they trigger agreement in the verb complex, see §2.1;
2. their sequential order, see §2.2; and
3. the selection of articles nominal arguments combine with.

2.1 Agreement

There are two morphemes within the verb complex (VC) that inflect like the pronouns for person and number: the object marker and the imperfective aspect marker. While the object marker indexes the primary object, hence its name, the imperfective aspect marker mostly agrees with the subject.

(6) \textit{Enaa [kahi hua nao nom]} \textit{VC.}  
\textit{1s.SBJ TAM paddle go 1s.IPF}  
‘I will paddle away.’ [Vae. 1.9E]

(7) \textit{Enaa [kahi asun u- nom]} \textit{VC- an.}  
\textit{1s.SBJ TAM kill 2s- 1s.IPF 2s.OBJ}  
‘I will kill you.’ [San. 1.107R]

(8) \textit{Nomana to [kahi asun va- mate- mate ri nom]} \textit{VC naa ori.}  
\textit{today REL TAM hit CAUS- RDP- dead 3p 1s.IPF 1s.SBJ 3p.OBJ}  
‘It’s today that I’ll kill them!’ [Eno. 1.53E]

Only if the subject is a third person and the primary object a first or second person does the object control the agreement of the imperfective aspect marker.

(9) \textit{be ori [von ama nom]} \textit{VC a- nam.}  
\textit{as.long.as 3p.SBJ pay 1pe 1pe.IPF 1pe- 1pe.OBJ}  
‘(We’ll do your work,) as long as they pay us.’ [Pur. 1.843R]

The table below shows the paradigms of pronouns, imperfective aspect markers and object markers. The object markers are clitics that, depending on the context, are attached to the preceding or the following element. When functioning as objects, the first person singular and the first person plural exclusive pronoun are indexed by the object marker \textit{a} or its variant \textit{ama} within the VC and, in addition, are marked by the prefix \textit{a-}. The fourth person object pronouns will be discussed below in §3.
Table 2: Pronouns, imperfective aspect marker and object marker

<table>
<thead>
<tr>
<th>Person</th>
<th>Pronoun</th>
<th>Imperfective aspect</th>
<th>Object marker</th>
<th>Object pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td>1s</td>
<td>enaa/naa</td>
<td>nom</td>
<td>a/ama</td>
<td>a-naa</td>
</tr>
<tr>
<td>1pe</td>
<td>enam/nam</td>
<td>nom</td>
<td>a/ama</td>
<td>a-nam</td>
</tr>
<tr>
<td>2s</td>
<td>ean/an</td>
<td>nom</td>
<td>u/vu</td>
<td>-an</td>
</tr>
<tr>
<td>2p</td>
<td>eam/am</td>
<td>nom</td>
<td>ame</td>
<td>-am</td>
</tr>
<tr>
<td>1pi</td>
<td>eara/ara</td>
<td>rara/ra</td>
<td>ara</td>
<td>-ara</td>
</tr>
<tr>
<td>3s</td>
<td>eove/eve/e/e/e/o</td>
<td>nana/na</td>
<td>–</td>
<td>eove/eve/e/e/-/-i</td>
</tr>
<tr>
<td>3p</td>
<td>eori/ori</td>
<td>rori/ri</td>
<td>ri</td>
<td>ori</td>
</tr>
<tr>
<td>4s</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>bona, bari</td>
</tr>
<tr>
<td>4p</td>
<td>–</td>
<td>–</td>
<td>ri</td>
<td>bari</td>
</tr>
</tbody>
</table>

2.2 Constituent order

The basic word order of Teop clauses is:

(10) TOPIC VERB COMPLEX OTHER ARGUMENTS

The topic position can be occupied by any of the three arguments, or be left empty if the topic is given by the context, or if a topic has not been established yet but is being introduced, as is the case at the beginning of a narrative. If the subject is the topic, the primary object directly follows the verb complex while the secondary object comes last. If one of the objects is the topic, the subject holds the first postverbal position. Table 3 presents the possible orders of arguments which are illustrated in examples (11) and (13).

Table 3: Sequential order of arguments

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>VC</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>unmarked</td>
<td>SBJ</td>
<td>OBJ1</td>
<td>OBJ2</td>
</tr>
<tr>
<td>marked</td>
<td>OBJ1</td>
<td>SBJ</td>
<td>OBJ2</td>
</tr>
<tr>
<td>marked</td>
<td>OBJ2</td>
<td>SBJ</td>
<td>OBJ1</td>
</tr>
</tbody>
</table>

(11) SBJ | VC | OBJ1 | OBJ2
[Me Toko] [paa hee] [bene Sookara] [bona overe].
[and:ART Toko] [TAM give] [OBJ.ART Sokara] [OBJ.ART coconut]
‘And Toko gave Sokara the coconut.’ [Sia. 1.68E]

(12) OBJ1 | VC | SBJ | OBJ2
[O piipi vai] [na dao ri- ] [ori] [bono toka-pis].
[ART.PL sore DEM] [TAM call 3p.IPF-] [3p.SBJ] [OBJ.ART.PL toka-pis]
‘(But on his two legs (he had) many sores.)
These sores, they call toka-pis.’ [Sia 1.18-19E]

3 The verb complex here does not contain an object marker although the primary object o piipi vai ‘these sores’ is plural. The object marker is frequently omitted with non-human NPs.
The fourth person pronoun

3.1 Two kinds of pronouns referring to non-speech-act participants

As already shown above in Table 2, Teop has two categories of personal pronouns for non-speech-act participants, henceforth called ‘outsiders’: the third person pronoun category comprising eove (3s) and eori (3p) (and their variants) and the fourth person pronoun category comprising bona (4s) and bari (4s/p). In transitive clauses the third person pronoun can function as the subject (14), or provided that the subject refers to a speech-act participant, as the object (15).

(14) Eove na rake nana bona.
3s.SBJ TAM want 3s.IPF 4s.OBJ
‘S/he wants it/him/her.’

(15) Enaa na rake nom e.
1s.SBJ TAM want 1s.IPF 3s.OBJ
‘I want it/him/her.’

However, if the subject is a third person as in (14), the object is expressed by a fourth person pronoun, e.g. bona. The same rule holds for the primary object of ditransitive clauses. If the subject refers to a speech-act participant, the primary object is expressed by a third person pronoun:

(16) Pesu ta veraka. Ean re tasu nao e bona.
pick NSPEC.ART ripe.breadfruit 2s.SBJ CNJ throw DIR 3s.OBJ 4s.OBJ
‘Pick a ripe breadfruit. Then you throw it at him.’
(lit. ‘Then you throw him it.’) [Jen. 1.69E]

With a third person subject, the primary object must be a fourth person pronoun:

(17) Me- ori paa dao bari bene Sirivana.
and 3p.SBJ TAM call 4s/p.OBJ OBJ.ART Sirivana
‘And they called her Sirivana.’ [MM 2.148R]

The fourth person pronoun only functions as an object and always presupposes the existence of a third person argument in a position that rank higher on the syntactic functions hierarchy: subject > primary object > secondary object (see §4.4).

---

4 The inalienable possessor, here the pronoun -e (3s.POSS), is linked to the possessed noun by a possessive marker that agrees in person and number with it, here -na- (3s).
Both the third person and the fourth person pronouns refer to outsiders, i.e. non-speech-act participants. Since the fourth person presupposes the existence of a third person, we call the referent of the third person pronoun a ‘first-order outsider’ and the referent of the fourth person a ‘second-order outsider’. Note that our use of the term ‘fourth person’ is different from its use in Amerindian linguistics, though it remotely resembles what Rice calls a ‘fourth person pronoun’ in her grammar of Slave, an Athapascan language, as this pronoun is only used for objects in clauses with a third person subject (Rice 1989:256, 628, 1196ff.).

### 3.2 Agreement patterns

The singular form of the fourth person pronoun *bona* (4s) refers to a singular second-order outsider as in (14) and (18):

(18) *Enaa paa sue wuru Aaron e iaa he vahuhu bona?*  
1s.SBJ TAM say earlier Aaron ART Mum CNJ give.birth 4s.OBJ  
‘Did I mention Aaron when Mum gave birth to him?’  [Sii. 3.263R]

However, the fourth person singular form *bona* (4s) is only used when the subject is also singular. If the subject is plural, the form *bari* (4s/p) is used, see (17) and (19).

(19) *Erau, me- ori paa dao bari bene Avelaua.*  
so and- 3p.SBJ TAM call 4s/p OBJ.ART Avelaua  
‘And so they called her Avelaua.’  [Sta. 1.40E]

In addition, *bari* refers to a plural second-order outsider irrespective of the number category of the subject:

(20) ... *bona bua beiko ... Evehe tetee na vaakisi ri bari.*  
OBJ.ART two child but Dad TAM not.allow 4p 4p.OBJ  
‘... the two children ... But Dad did not allow them (to go).’  [Sii. 3.294–296R]

That *bari* in (20) refers to a plural participant can not only be inferred from the context (*bona bua beiko*), but is also indicated by the plural object marker *ri*. This object marker is missing in (17) and (19) where *bari* refers to a single person.

### Table 4: The syntactic functions of the fourth person pronouns

<table>
<thead>
<tr>
<th>Person of subject (SBJ)</th>
<th>Person of primary object (OBJ1)</th>
<th>Person of secondary object (OBJ2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1/2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 5: Agreement patterns of the fourth person pronoun

<table>
<thead>
<tr>
<th>Primary object</th>
<th>Secondary object</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th person SG</td>
<td>4th person PL</td>
</tr>
<tr>
<td><em>bona</em> (4s)</td>
<td><em>bari</em> (4p)</td>
</tr>
<tr>
<td><em>bari</em> (4s/p)</td>
<td><em>bari</em> (4p)</td>
</tr>
<tr>
<td>4th person SG</td>
<td>4th person PL</td>
</tr>
<tr>
<td><em>bona</em> (4s)</td>
<td><em>bari</em> (4p)</td>
</tr>
</tbody>
</table>
The agreement rule of *bona/bari* is only applied when these fourth person pronouns function as primary objects, as is illustrated by the following two examples:

(21) SBJ OBJ OBJ OBJ 
... o- re paa mosi suku maa bari bona ...
3s/p CNJ TAM cut following DIR 4s/p 4s
‘... so that they cut it (the tree) as he wants it ...
(lit. ‘so that they cut following him it’) [Tah. 2.22R]

The incorporation of the preposition *suku* ‘following, according to, because of’ makes the verb complex ditransitive. While *mosi* ‘cut’ takes the patient as its primary object, the primary object of *mosi suku* refers to the person who is followed or according to whom the cutting is done. Thus the primary object *bari* means ‘him’ here. It refers to a single person (as is known from the context), but has a plural form because the subject pronoun refers to more than one person. In contrast, the second pronoun, *bona* ‘it’, which refers to the patient, does not show this kind of agreement. However, when *mosi* ‘cut’ is used by itself in a transitive clause with a 3PL subject, the pronominal patient object is *bari*:

(22) ... o- re paa mosi bari ...
3p CNJ TAM cut 4s/p
‘... then they cut it (the tree) ...’ [Tah. 2.25R]

As above, in example (19), *bari* unambiguously refers to a singular participant, namely the tree that is used for carving a canoe.

### 4 Case marking by articles

The articles can be classified into three classes: specific articles, non-specific articles and the partitive article. The specific articles are further divided into basic and object articles (abbreviated ART and OBJ.ART). Both the specific and the non-specific articles distinguish between three noun classes and singular and plural (Mosel and Spriggs 1999b). For the coding of arguments, only the distinction between basic and object articles is relevant.

<table>
<thead>
<tr>
<th>Table 6: Basic and object articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun classes</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>basic article</strong></td>
</tr>
<tr>
<td>e-class</td>
</tr>
<tr>
<td>singular</td>
</tr>
<tr>
<td>ere, o</td>
</tr>
<tr>
<td>plural</td>
</tr>
<tr>
<td>bone, bene, bere, benere, bona</td>
</tr>
<tr>
<td><strong>object article</strong></td>
</tr>
<tr>
<td>a-class</td>
</tr>
<tr>
<td>singular</td>
</tr>
<tr>
<td>o</td>
</tr>
<tr>
<td>plural</td>
</tr>
<tr>
<td>bona, bono</td>
</tr>
<tr>
<td>o-class</td>
</tr>
<tr>
<td>singular</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>plural</td>
</tr>
<tr>
<td>bono, bona</td>
</tr>
</tbody>
</table>

The object articles are related to the demonstratives *bene* ‘this’ and *bona* ‘that’, but in contrast to the object articles, these demonstratives do not distinguish noun class and number, and they always follow the noun phrase nucleus. In the spoken language the object article is sometimes followed by a basic article, often with a little pause in between so that this could be interpreted as a new start:
(23)  *bona a si aba*

OBJ.ART ART DIM person

‘the little person’  [Iar. 2.65R]

However, this construction is also found in the transcriptions and edited versions of transcriptions done by native speakers. For example:

(24)  *bona a are*

OBJ.ART ART betel.nut

‘the betel nut’  [Nan. 2.7E]

Therefore it does not seem to be justified to dismiss such instances as a speech error or a hesitation phenomenon. One possible solution would be not to classify the object articles as articles but as prenominal demonstratives and say that the article is optional after a prenominal demonstrative. This interpretation does not have any impact on the analysis of the function of these words.

4.1 Article selection rules

In interaction with word order and cross-referencing pronominals, basic and object articles contribute to the coding of the syntactic functions of nominal arguments according to the following rules that are called article selection rules:

1. Topics always hold the first position in the clause and take the basic article irrespective of their syntactic function, see examples (11)–(13).
2. Irrespective of their position, subjects always take the basic article, see examples (1), (26) and (27).
3. Non-topical primary objects take the basic article when the subject of the clause is first or second person (25), otherwise they take the object article, see example (26).
4. Non-topical secondary objects take the basic article only when both the subject and the primary object refer to the speech-act participants, see examples (28) and (30). Otherwise they take the object article, see examples (29) and (31).

Table 7: Nominal arguments marked by basic and object article (ART and OBJ.ART)

<table>
<thead>
<tr>
<th>Subject in topic position</th>
<th>Non-topical primary object</th>
<th>Non-topical secondary object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st/2nd person pronoun</td>
<td>1st/2nd person pronoun</td>
<td>3rd person pronoun/ART N</td>
</tr>
<tr>
<td>1st/2nd person pronoun</td>
<td>3rd person pronoun/ART N</td>
<td>OBJ.ART N</td>
</tr>
<tr>
<td>3rd person pronoun/ART N</td>
<td>1st/2nd person pronoun</td>
<td>OBJ.ART N</td>
</tr>
<tr>
<td>3rd person pronoun/ART N</td>
<td>OBJ.ART N</td>
<td>OBJ.ART N</td>
</tr>
</tbody>
</table>

4.2 Article selection in transitive clauses

Examples (25)–(27) below illustrate the article selection rules; the underlined phrases are arguments that refer to outsiders.

Compare the object marking in the following transitive clauses:
(25)  
\[ \text{Enaa paa dee- ma-u e guu.} \]
\[ \text{1s.SBJ TAM carry- DIR-IM ART pig}\]
\[ \text{‘I have brought a pig.’ } [\text{Mat. 1.78E}] \]

(26)  
\[ \text{A beiko te- naa paa asun- u bene guu.} \]
\[ \text{ART child PREP- 1s TAM kill- IM OBJ.ART pig}\]
\[ \text{‘My child has killed the pig.’ } [\text{Aro. 15.134R}] \]

If the object is topicalised, it is still possible to distinguish subject and object:

(27)  
\[ \text{E guu paa asun- u a beiko te- naa.} \]
\[ \text{ART pig TAM kill- IM ART child PREP- 1s}\]
\[ \text{‘The pig has been killed by my child.’}\]

This clause cannot mean ‘The pig has killed my child.’ because then ‘my child’ would be the object and marked by the object article: \[ \text{E guu paa asunu bona beiko tenaa.}\]

4.3 Article selection in ditransitive clauses

Article selection in ditransitive clauses is nicely illustrated in the following story about a one-legged boy called Gaagin and his friend Biroo, a lizard. When Biroo wants to go fishing with Gaagin, Gaagin says that he can’t walk to the reef because he has only one leg, but Biroo gets him the leg of his grandmother and they go fishing at the reef. There they fight over some fish and Biroo takes his grandmother’s leg from Gaagin. Another day Biroo wants to go with Gaagin to the reef again, but Gaagin says:

(28)  
\[ \text{Ean [na gono kahi vahabana ama-] [naa]} \]
\[ \text{2s.SBJ TAM take from again 1s 1s.OBJ}\]
\[ \text{you took again from me}\]
\[ \text{OBJ2}\]
\[ [a moo- n- e subu- m- an]. \]
\[ \text{ART leg 3s ART grandmothe- 2s -2s.POSS}\]
\[ \text{the leg of your grandmother}\]
\[ \text{‘You took again your grandmother’s leg from me.’ } [\text{Aro 4.93R}] \]

Here the secondary object takes the basic article because the subject and the primary object refer to speech-act participants so that the secondary object is a first-order outsider.

The following section of the story shows the use of the object article with the secondary object. The context is nearly the same, the only difference being that the primary object is a third person, i.e. a first-order outsider, so that the secondary object becomes a second-order outsider. In the story, Biroo again gets the leg from his grandmother and fixes it onto Gaagin, but as before, they fight over some fish and Biroo pulls off the leg from Gaagin and walks home, leaving Gaagin alone on the reef. When Biroo comes to the village, Gaagin’s mother asks for her son, and Biroo says that he already went home. Gaagin’s mother then replies:

(29)  
\[ \text{Ean na piku nom,}\]
\[ \text{2s.SBJ TAM lie 2s.IPF}\]
\[ \text{paa gono kahi bana ma-u e Gaagin}\]
\[ \text{TAM get from again DIR-IM ART Gaagin}\]
Ulrike Mosel

[bona moo- n- e subu- m- an].
OBJ.ART leg- 3s- ART grandmother 2s- 2s.Poss
‘You are lying, (you) have again taken your grandmother’s leg from Gaagin.’
[Aro 4.131-132R]

Being the primary object and a first-order outsider argument e Gaagin is marked by the basic article, while the secondary object bona moone subuman ‘your grandmothers’ leg’ is a second-order outsider and marked by the object article.

In another story, we find two clauses of nearly the same content that illustrate the use of the fourth person object pronoun and the alternation of the basic and the object article with the secondary object. In this story a giant has taught the people how to prepare ariono vines for making fishing nets. When they have not done it properly, he says:

(30)  Enaa pasi vaasusu avameam⁵ a meha taba.
      1s.SBJ TAM teach 2p.OBJ.IM ART other thing
‘I’ll teach you another thing.’   [Sii. 6.45R]

The story continues:

(31)  Me paa vaasusu ri bari bona meha taba.
      and TAM teach 4p 4p OBJ.ART other thing
‘And (he) taught them another thing.’   [Sii. 6.46R]

The implied subject of the second clause (‘he’) is a third person, i.e. a first-order outsider, and consequently the primary object ‘them’ is expressed by the fourth person pronoun and the secondary object is marked by the object article.

The following two elicited examples show that the secondary object is also marked by the object article when only one of the other arguments — the subject or the primary object — is a third person:

(32)  Eori paa vaasusu ara- ara⁶ bona meha taba.
      3p.SBJ TAM teach 1pi- 1pi.OBJ OBJ.ART other thing
‘They taught us another thing.’

(33)  Enaa paa vaasusu ri- a abana bona meha taba.
      1s.SBJ TAM teach 3p- ART men OBJ.ART other thing
‘I taught the men another thing.’

The word form avameam is a fusion of ameam, which consists of the object marker ame- (2p) and the pronoun -am (2p), and the so-called immediateness marker vu/u, a tense/aspect marker that holds the position after the object marker within the verb complex.

The word form araara consists of the 1pi object marker ara and the 1pi pronoun ara.

The wordform araraara consists of the 1pi object marker ara, the 1pi imperfective aspect marker ra and the 1pi pronoun ara. The imperfective aspect marker cross-references the object because the object rank higher on the person hierarchy than the subject.
4.4 Speech-act roles

As illustrated above, the coding of non-topical objects can only be explained if one considers the argument structure of the clause and, in addition, the speech-act roles of the arguments:

- the speech-act participants (1st and 2nd person)
- first-order outsiders (3rd person)
- second-order outsiders (4th person)

In clauses that contain only one outsider argument as in (25) and (30), this argument is a first-order outsider irrespective of its syntactic function. Second-order outsiders imply the existence of a first-order outsider. The former is distinguished from the latter in that it figures lower on the syntactic functions hierarchy:

(35) subject > primary object > secondary object.

Accordingly, the following four rules hold:

1. If in a transitive clause the subject refers to a speech-act participant and the object to an outsider, the object is a first-order outsider argument, see (25).
2. If in a ditransitive clause the subject refers to a speech-act participant and the objects to outsiders, the primary object is a first-order outsider argument and the secondary object a second-order outsider argument, see (29).
3. If the subject of a ditransitive clause is an outsider argument, it is a first-order outsider argument, while all other outsider arguments are assigned the status of a second-order outsider argument, see (11)–(13), and (31).
4. The secondary object is only a first-order outsider argument when both the subject and the primary object refer to speech-act participants, see (28) and (30).

All three types of core argument — the subject, the primary object and the secondary object — can be topicalised and then are marked by the basic article irrespective of their syntactic function. The argument structure of the clause, however, remains the same, as is shown by the cross-referencing morphology in the verb complex and the order of arguments following the verb complex. Topicalisation of an argument also does not affect the speech-act role structure of the clause as the following example illustrates:

(36) O vioga me na dao- dao bata ra- ara
   ART vioga also TAM RDP- call at.the.same.time 1pi.IPF- 1pi.SBJ
   bona are te Magaru.
   OBJ.ART betelnut of:ART Earthquake
   ‘The vioga palm, we also call Earthquake’s betelnut palm.’ [Val. 2.61E]

The verb dao ‘call (s.o. s.th.)’ is ditransitive. Its primary object denotes the recipient (o vioga) and its secondary object the name (bona are te Magaru). As the primary object (o vioga) is topicalised, the secondary object directly follows the subject (ara ‘1pi’), but it is still a second-order outsider argument and consequently marked by the object article.
4.5 Summary: the article selection rules

The preceding sections show that the selection of the article follows the following rules:
– Subjects always take the basic article.
– Topics always take the basic article.
– Non-topical objects take the basic article if they refer to first-order outsiders.
– Non-topical objects take the object article if they refer to second-order outsiders.

This means that the case marking of arguments by articles is determined by three parameters:
– their pragmatic role as a topic or non-topic;
– their syntactic function as a subject, primary object or secondary object; and
– their category of person, i.e. their speech-act role as a speech-act participant, a first-order outsider or a second-order outsider.

5 Conclusion: the interaction of pragmatics, semantics and syntax

While most languages only distinguish between speech-act participants, i.e. the first and the second person, and one category of outsiders, i.e. the third person, Teop formally distinguishes between two kinds of outsiders (that is, non-speech-act participants), namely the third and the fourth person. This distinction manifests itself in the paradigm of pronouns and in the selection of articles with nominal arguments.

If a clause has more than one outsider argument, the argument that ranks higher on the syntactic function hierarchy (subject ≥ object 1 ≥ object 2) has the status of the first-order outsider, i.e. the third person, whereas the other argument or arguments are second-order outsiders or fourth persons. This means that outsider subjects are always third persons.

The category of the fourth person is only found with objects and always presupposes another outsider argument that ranks higher on the syntactic functions hierarchy. Accordingly, a primary object is a fourth person when it refers to an outsider and when the subject is already an outsider, i.e. a first-order outsider or third person. Similarly, a secondary object is a fourth person when either the subject or the primary object or both are already outsiders. Conversely, a third person primary object only occurs in clauses with a first or second person subject, and a third person secondary object presupposes that both the subject and the primary object refer to speech-act participants.

The syntactic functions of arguments are encoded by an intricate interaction of four coding strategies:

1. the sequential order of arguments in non-topical position after the VC;
2. the object marker that cross-references the primary object;
3. the imperfective aspect marker that cross-references the subject or the object; and
4. the selection of the basic article or the object article with nominal arguments or the corresponding third and fourth person pronouns.

These are presented in Table 8.
### Table 8: Encoding of arguments

<table>
<thead>
<tr>
<th>Clause initial position if topic</th>
<th>OBJ1 if topic</th>
<th>OBJ2 if topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential order in non-topical position after VC</td>
<td>SBJ &gt; OBJ1</td>
<td>SBJ &gt; OBJ1</td>
</tr>
<tr>
<td></td>
<td>SBJ &gt; OBJ2</td>
<td>OBJ1 &gt; OBJ2</td>
</tr>
<tr>
<td>Cross-referenced by object marker</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Cross-referenced by imperf. aspect marker</td>
<td>if equal to or higher than object 1 on person hierarchy</td>
<td>if higher than subject on person hierarchy</td>
</tr>
<tr>
<td>Basic article</td>
<td>always</td>
<td>if topic or third person</td>
</tr>
<tr>
<td>Object article</td>
<td>–</td>
<td>if fourth person</td>
</tr>
</tbody>
</table>

With the exception of the object marker, all coding strategies have more than one function. They not only express the syntactic function of arguments, but also relate to the topicality and speech-act roles of arguments. The discourse function of topic is independent of the categories of speech-act role and syntactic function as any kind of argument can function as the topic.

The above analysis also showed that to fully understand the function of constituent order and the selection of articles, ditransitive clauses cannot be treated as a marginal type of clause, but have to be fully integrated into the descriptive framework.

### References


—— forthcoming, *Ditransitivity in Teop*.


Measures and morphemes: a functional approach to Hawaiian accent

ALBERT J. SCHÜTZ

1 Introduction

Many Hawaiian songs come to a close by announcing that the theme — often a person or a place being honoured — will be repeated:

Ha‘ina ‘ia mai (ana) ka puana. ‘Let the summary refrain be told.’

Before I knew the meaning of this phrase, I thought that the last two words were kapu ana. This purely auditory impression, without reference to meaning, lay dormant in my mind for some years, not to be drawn upon until I became interested in Hawaiian prosody. Then it became clear that my misconception about word boundaries in this phrase was a result of the complicated relationships among measures (accent units), content morphemes, and function morphemes. This paper examines some of those relationships, using phrases — not just words — as data.
2 Previous treatments of Hawaiian accent

In the published material on Hawaiian, accent has not been treated in depth. But some early works outline the main principles (e.g. Schütz 1978, 1986). Applying the concept of the measure (a unit defined by accent; see §4) to Hawaiian was a natural outcome of my work on Fijian, whose accent system is nearly identical except that long vowels and diphthongs obligatorily shorten in the penultimate position of the measure. A sketch (Schütz 1978) describes accent in these two languages, emphasising the point that it is predictable only at the level of the measure, and that longer words can be best described as sequences of measures. The 1986 work is a short explanation of what underlies the marking system in the headwords in the revised Pukui-Elbert dictionary.

This major revision of the dictionary’s treatment of accent was necessary because of the incorrect statements in previous editions. For example, the 1961 edition contained the following rules (p. xxix):

**Stress (or Accent)**

On all vowels marked with macrons: ā, ē, ī, Ħ, ū.

Otherwise on the next-to-last syllable and alternating preceding syllables of words, except that words containing five syllables without macrons are stressed on the first and fourth syllables: hāle, makāʻu, hōlōhōlo, ʻelemakāle [‘house’, ‘fear’, ‘walk, ride’, ‘old man’].

Every word or particle separated by spaces is stressed except clitics …

The first problem with this description is that it does not mention that a diphthong functions as a syllable peak, with the first vowel of the pair accentuated no matter what its position in a word or accent unit. Next, as pointed out in Schütz 1978:143, many five-syllable words, such as makūahīne ‘mother, aunt’ do not follow the second rule. In spite of the revision in the unabridged dictionary, the most recent edition of *Pocket Hawaiian dictionary* (Pukui, Elbert et al. 1992) still contains the incorrect description.

Unfortunately, the alternating-syllable portion of the statement seems to have been accepted without question, despite evidence to the contrary. For example, Hayes (1995:181) listed Hawaiian among ‘Languages with Fijian-like stress’. Therefore, it is necessary to approach his view of Hawaiian accent through his description of Fijian accent (pp. 142–149).

Although in the latter part of this treatment Hayes introduced an analysis based on moraic trochees, his account of the ‘stress pattern of Fijian’ (p.142) echoes most of the points in the description of Hawaiian quoted above:

a. If the final syllable is light, main stress falls on the penult.

b. If the final syllable is heavy, main stress falls on the final syllable.

c. Secondary stress falls on remaining heavy syllables, and on every other light syllable before another stress, counting from right to left.

For both Hawaiian and Fijian, portions of the descriptions are correct. For example, Hayes’s statements (a), (b), and the first half of (c) are valid. The incorrect portion is the phrase in (c) referring to alternating syllables. In words consisting of only light syllables, any combination of a disyllable followed by a trisyllable (e.g. Hawaiian pūlelehua ‘butterfly’; hōnekakāla ‘honesuckle’) results in a succession of two light syllables, which disrupts the pattern of alternating syllables. Although such forms can be either native or borrowed, the latter provide a convenient list of so-called exceptions, prompting Hayes to
explain the discrepancy by suggesting that some loanwords ‘indicate that secondary stress can sometimes be assigned from left to right’ (1995:144). 3

Terms such as ‘every other light syllable’ and ‘counting from right to left’ make it sound as though the primary data underlying these rules are written word lists (with the accented syllables already marked, of course). Is it realistic to suggest that the accents in hùmuhùmunùkunùkuâpuâ‘a ‘trigger fish’ can be accounted for by counting backwards from the end of the word, while those in kilokalâme ‘kilogram’ require counting from the beginning of the word?

As words combine into phrases and sentences, counting syllables (from either end) seems even less likely. This leads to the question: how do speakers and hearers handle accent in phrases and sentences?

For the present paper, the crucial word in that question is ‘hearer.’ Admittedly, speech involves both the speaker and the hearer, but here, I oversimplify the matter by assuming that while speakers know what they’re saying, hearers must rely on auditory clues to interpret the string of speech into recognisable units.

3 Phonology for the hearer

The phrasing in the two descriptions above, especially the part about alternate syllables and counting from either direction, reminds me of a term I heard often from Charles F. Hockett when I was his student: marble-slab grammar. He explained it in an article entitled ‘Grammar for the hearer’ (1961:220). First, he introduced the main topic:

This paper deals, in a preliminary way, with the problem of grammatical design from the vantage point of a hearer: that is, of a person who knows a language and who, for the moment, is silently listening to someone else speak.

Next, he sketched the common view of grammatical analysis at that time:

… The grammarian can view a sentence as an enduring structure, to be scanned at leisure and repeatedly, and as easily from right to left … as from left to right. He can do this because he deals not directly with a sentence, but only with a representation thereof, spread out before him like a cadaver on a marble slab, to be dissected at his convenience.

Finally, he stated his objection to this methodology:

The hearer has none of these advantages. He is exposed to an utterance just once, and is forced to register its ingredients in just the temporal sequence in which they reach him.

I would like to apply Hockett’s theme to Hawaiian prosody, changing his title to ‘Phonology for the hearer.’ 4 And as he did in his introductory sentence, I add the phrase ‘in a preliminary way’; the present work is only a beginning.

I suggest that the main reason for the failure of rules such as the ones cited for Fijian is that the logic behind them is backwards. That is, it ignores the main function of accent in

---

3 In my analysis, the distinction between primary and secondary accent is a feature of the phonological phrase, and each word pronounced in isolation is a phonological phrase. Unless there is contrastive accent, the last measure in the word (phrase) is emphasised to mark its end. In this paper, so-called ‘secondary accents’ are marked only for the reader’s convenience.

4 This is not a new position for me. In 1978 (p.144) I wrote: ‘But any approach to accent from the speaker’s point of view seems difficult now. I prefer to examine accent from the hearer’s end of the communication process.’
Fijian and Hawaiian, which is neither regular nor contrastive (i.e. neither predictable nor phonemic), but demarcative. In other words, just as intonation guides us (but not always infallibly) to two grammatical units — phrases and sentences — accent helps the hearer sort out strings of otherwise similar syllables into units that correspond (again, not infallibly) to morphemes. In other words, it is accent that guides the hearer to the morphology, not the other way around.\(^5\) Both approaches are important, but the former seems to have been neglected in descriptions of Oceanic languages.\(^6\)

Another shortcoming of the previous treatments is that they have dealt with accent mainly at the word level. The present study attempts to examine measures not just in Hawaiian words, but also in phrases and sentences. The first step toward this goal is to list the different types of measures that occur.

### 4 Measure types

As in my previous studies on this topic, I posit a fundamental unit of accent — the ‘measure’.\(^7\) To find all possible measure types, I examined words that contain only one accented syllable.\(^8\) Table 1 shows these types, with examples. In the column labelled ‘Shape’, \(V\) = vowel, \((C)\) = optional consonant, \(D\) = diphthong, and a macron indicates length. The vowels are \(a, e, i, o, u, ā, ē, ī, ō, ū\). Diphthongs are \(ai, au, ae, ao, ei, eu, oi, ou,\) and \(iu\). Long diphthongs are fewer and less frequent in occurrence. The only certain ones are \(āi\) and \(āu\), but \(āe, ēi, ōu, āo,\) and \(ōi\) are also possibilities.\(^9\) This treatment expands my earlier description in Pukui and Elbert\(^{10}\) mentioned above (see Schütz 1986:xvii–xviii).

---

\(^5\) This does not mean that measures and morphemes always share the same boundaries. See §5 for a discussion of the prosodic behaviour of different combinations of morphemes.

\(^6\) As examples of descriptions that rely partly on morpheme boundaries, see Biggs 1978:700 and Feldman 1978:134.

\(^7\) Although I have used the term ‘measure’ in a number of works since 1976, readers unfamiliar with those works need to know how this unit differs from a foot. Hayes (1995:143–144) described the difference as follows: ‘This [Hayes’ treatment of Fijian] is to a large extent a translation of the analyses in Scott 1948 and Schütz 1985. These scholars parse the word into ‘groups’ (Scott) or ‘measures’ (Schütz), which are the same as moraic trochees, except that they may optionally incorporate a stray light syllable at the beginning of the metrical unit … There appear to be no data within Fijian that could decide between these two views of metrical constituency, though only moraic trochees would generalise to other languages.’ This explanation needs two minor corrections. First, although Scott confined his ‘groups’ to words, I have extended the concept to apply to phrases and sentences as well. Next, as my studies of prosody in Hawaiian, Māori, Tongan, and Samoan have shown, the analysis applies to those languages as well; although a trochaic view does not explain accent patterns in perhaps all Nuclear Polynesian and closely related languages. As for Hayes’s short comments on Hawaiian and Tongan, he may have been hampered by the inadequate descriptions available (e.g. Feldman 1978, Elbert and Pukui 1979).

\(^8\) This is an expansion and refinement of my earlier descriptions of Hawaiian measures (Schütz 1978, 1986).

\(^9\) I included the second set in Schütz 1981. However, most of the examples were not interpreted that way in Pukui and Elbert 1986. Dialectal or idiolectal variation, in addition to possible mistranscriptions by Elbert, may account for the discrepancy. Still, the measure type exists, no matter how many specific members there are. See the entry for \(kēua\) in Schütz, Kanada and Cook 2005.

\(^{10}\) In the first printing of Pukui and Elbert 1986, the accent over long vowels was omitted. In later printings, this error was corrected.
In Table 1, accented syllables are marked, although once measure divisions are indicated, the location of the accented syllable is predictable according to three principles. Accent occurs on:

1. a syllable with a long vowel as the nucleus,
2. a syllable with a diphthong (short or long) as the nucleus, or
3. the penultimate syllable in a measure made up of light syllables.

The division into even and uneven measures is crucial to this analysis:

An even measure consists of:

- an accented heavy or superheavy syllable (i.e. a long diphthong), or
- a disyllable consisting of light syllables, with the first syllable accented.

An uneven measure consists of an even measure preceded by a light unaccented syllable.

Thus, as can be seen in Table 1, these fourteen specific types of measures can be grouped into two main types: even and uneven. Each uneven measure consists of any of the seven types of even measure preceded by a light unaccented syllable. They can also be divided into monosyllables, disyllables, and trisyllables. Finally, they can be represented by the following notation, in which \( \tilde{V} \) represents a light unaccented vowel:

\[ \text{In this group, the second consonant is not always optional. For example, deleting the second consonant from the disyllable } luna \text{ 'above' yields } lua \text{ 'hole', another disyllable, but the same operation on the disyllable } manu \text{ 'bird' yields } mau \text{ 'eternal', a monosyllable. Also if the vowels are identical, the result is a long vowel. Strictly speaking, each vowel pair should be listed separately, but for the sake of simplicity this has not been done.} \]
an accented heavy or superheavy syllable

\{ (C)V \} +

OR

a disyllable, with first syllable (of any weight) accented, and the second syllable light and unaccented

5 An analysis of the data

The data for this paper are the first six brief stories from *The Hawaiian sentence book* and its accompanying tape (Snakenberg 1988:2–7). The stories, with interlinear transcriptions and measure notation added, appear in the Appendix. Although read in a rather formal style, these stories still show some features of more casual speech, such as occasional shortening of long vowels and variation between diphthongs and vowel sequences. Such changes are identified in the footnotes. Perhaps this kind of text is not ideal, but I consider it a starting point: it is easier to derive casual speech from formal speech than the other way around. The conventions used in this section are those given in the Appendix.

In the data examined, as discussed in the previous section, we can find a number of different types of measures, based on their morphological content. The following list shows these types, the number of instances in the data, examples, and the location (see the Appendix) of each example.

1. M 50 \[ki'i\] (1.1)

   This is a one-to-one match between measure and morpheme — that is, one measure consists of one content form.

2. m 32 \[kē\] (1.1), \[ua\] (2.4), \[paha\] (4.2)

   Some common function forms, such as \[ua\] (aspect), \[nā\] (plural), \[paha\] ‘perhaps’, and \[ia\] (third person singular), consist of an even measure. These are also one-to-one matches.

3. mM 18 \[he ki'i\] (1.1)

   Most examples of this type consist of certain function forms that are a short syllable in length, followed by a content morpheme whose form is an even measure. Because of the high frequency of the articles \[he\] and \[kalke\] and several prepositions, as well as content morphemes that consist of an even measure, this type is common. As pointed out by Kenneth W. Cook (pers. comm., 1 August 2006), in the 1820s, when Hawaiians were learning to write, some had a tendency to write such phrases as \[ka puke\] ‘the book’ as a single word: \[kapuke\].

4. mM 10 \[he ka\]naka (4.2)

   This example shows the result of a light syllable preceding an uneven measure: the resultant realignment produces a mismatch between measure and morpheme. The first measure consists of a function morpheme and part of a content morpheme; the next consists of only part of a content morpheme. In the example, the article \[he\], normally unspecified as to level of accent, is now the peak syllable of a measure and therefore
Measures and morphemes: a functional approach to Hawaiian accent

accented. However, because it is a function morpheme, its accent level is somewhat reduced. In the content morpheme, the accented syllable retains its accent.

5. \[37 \text{he ka} \{naka\} (4.2)\]

Example #4, repeated here, also shows that part of a morpheme — naka — can serve as a separate measure. Note that the realignment at the beginning of the phrase produces no change in the accent pattern of the base form kanaka.

6. \[19 \{'o ka\} (1.4), \{i ka\} (1.4)\]

Certain combinations of function forms that consist of light syllables are so frequent that the construction has sometimes been interpreted as one word rather than two. An example is me he \{mehe\}\(^{12}\) (see the entry mehe, me he in Schütz, Kanada and Cook 2005).

7. \[15 \{'o ia\} (2.4), \{o nā\} (6.1)\]

Similarly, the combination of two function forms — one a light syllable and the other heavy — has also sometimes been interpreted as one word, other times as two. ‘O ia (‘oia?) is an example. Phonologically, o and i unite as a diphthong, but morphologically, each belongs to a separate unit.

8. \[2 \{i ko-na\} (4.3)\]

This example shows the maximum number of morphemes contained in one measure: three.

9. \[1 \{hele-na\} (5.3)\]

This is an example of a content morpheme followed by a monosyllabic suffix, resulting in what is sometimes called a ‘shift in accent’, but is actually a movement of syllables under the accent contour. If there are no heavy syllables involved, this type results in the peak syllable in the base form becoming unaccented, and vice versa. See #11 and the discussion in §6.2.4.

10. \[1 \{ikaika\} (3.2)\]

This notation seems to indicate a contradiction: two content morphemes, which must be at least a heavy syllable in length, serving as one measure. However, the example, ikaika ‘strong’, is the result of (1) reduplication, and (2) resyllabification of the resultant vowel combination \(a + i\). Similarly, a combination of like vowels formed through reduplication can fuse into a long vowel, resulting in one measure rather than two, e.g. aka + aka → akāka.

11. \[1 \text{maka-‘āi} \{na-na\} (4.1)\]

Here, the second syllable of a morpheme — ‘āina — combines with a following light syllable to form a measure. The example is similar to #9 in that the final -na moves the previous syllable (unaccented in the base form) to the peak position in the measure. The peak syllable of the base morpheme, ‘āi, retains its accent since it is a long diphthong.

\(^{12}\) Analysts disagree whether it is one word of two.
12. mmM 3 [o ke a]kua (1.3)
13. mm 4 [lāua] (with resyllabification) (5.3)

The resyllabification reduces mm to one measure. Without such a change, it would be two measures.

14. mmm 1 ['o ke-ia] (1.3)

Diphthong shortening allows these three types to fit into one measure.

6 Tentative conclusions

The hypothesis underlying this study is that morphology cannot properly be used to predict accent, but instead, that accent helps guide the hearer to the morphological structure. How well do the data support the hypothesis? The answer lies in how we interpret them. There are two possible ways:

1. We can examine the degree to which measure and morpheme boundaries coincide (see the examples in §6.1), or
2. We can examine the degree to which the measure and morpheme peaks coincide (see the examples in §6.2).

We now analyse the data from each of these points of view.

6.1 Boundaries

There are 194 measures in the data. Of these, 82 (42%) coincide with morphemes. (For example, in 1 Akua in the Appendix, the morphemes kē- and ia (1.1, 1.2), akua (1.2), ki‘i (1.3, 1.4), wā (1.4), kahiko (1.4), and mana (1.4) correspond to measures, but there are more mismatches than matches.) This figure shows that measure boundaries do not significantly coincide with morpheme boundaries. As examples of mismatches, note the many instances of measures that consist of mM, mM, M, mm, and mm.

6.2 Peaks

To apply the second methodology, we ask: How often does the peak syllable of a measure coincide with the/a peak syllable of a morpheme? Of the 194 measures in the data, 164 (84.5%) show this kind of match, exemplified by all the measures in Appendix §1.1: [he ki‘i], [lā]’au], and [kē-][ia]. This is a more satisfactory figure.

But we still need to consider the mismatches. What is the nature of the 30 examples that do not match? They fall into several groups:

---

13 ‘Morpheme peak’ refers to the accented syllable in a stem (i.e., a content morpheme without affixes, or a function morpheme a measure in length (e.g. ho‘o- ‘CAUS’ or ana ‘NR’). Morphemes longer than one measure have more than one peak.

14 See Schütz 1978 for similar ideas. This approach seems instinctively better: it is easier to identify peaks than borders.
6.2.1 Combinations of function morphemes

Some measures are composed of two or three monosyllabic function morphemes. Given that each has no intrinsic accent level of its own,\(^\text{15}\) it is, in a sense, an accentual chameleon, taking its accent level from its position in the newly formed measure. For example, in Appendix sentence 1.4, we find the following measures: [i ka], ['o ke], and [o ka], each of which is a combination of two monosyllabic function forms, the first of which is accented — but less strongly than the peak syllable of a content morpheme.

Some such combinations of function forms are very common, and may lead hearers to interpret them as words. As the following section shows, certain types of combinations are especially likely to be heard as words.

6.2.2 Resyllabification

If, when two forms are combined, the second begins with a vowel, the resultant vowel combination may resyllabify and form a diphthong if it is one of the set listed in §4. This process produces another discrepancy between morphology and phonology. For example, in the phrase ‘o + ia ‘he, she’, sound and meaning pull in opposite directions, making it hard to decide whether this form is one word or two. In terms of sound, o and i resyllabify, and the accent pattern (‘óia) suggests that the phrase is one word. However, the division of the morphemes as ‘o ia argues for interpreting it as two words. Following the latter argument, in 1978, the official committee on spelling (‘Ahahui ‘Ōlelo Hawai‘i) decided that it should be written ‘o ia, probably since ‘o is written as a separate word before names, e.g. ‘o Pua, ‘o Keola.

Another example is the expression, widely used in local English as well, that means ‘toward inland, toward the mountains’. In terms of meaning, it is made up of two parts: ma (directional marker), and uka ‘inland, interior’. The fact that ma is used in the same way in many other phrases (e.g. ma Honolulu), is an argument for writing this expression as two separate words: ma uka. On the other hand, pronunciation supports writing it as one word, since the a and u form a diphthong. Thus, the ‘phrase’ is usually pronounced as a unit: máuka, not ma úka.

6.2.3 Morphemes longer than a measure

Some morphemes stretch over more than one measure. For example, in Appendix sentence 1.1, [lā]/[‘au] ‘wooden’ comprises one morpheme, but has two accent peaks. Perhaps each of these forms was once a morpheme, but related words in the Hawaiian dictionary and the word’s Proto Eastern Polynesian etymology do not support this explanation. Other examples are [‘ai]/[lona] ‘symbol’, [mahī]/[ole] ‘helmet’, [‘ā]/[kau] ‘right’, [lō]/[ihi] ‘long’, [kau]/[ā] ‘outcast’, [kā]/[kau] ‘tattoo’, and [‘ī]/[lio] ‘dog’.\(^\text{16}\)

Still, even though in these longer forms there is no one-to-one match between measures and morphemes, the peaks still match. In other words, the accents of the morphemes in context match those of the morphemes in citation form.

---

\(^{15}\) Some analysts maintain that such forms are unaccented in any position (for example, see Arms 1987:115; I disagree.

\(^{16}\) This word appears in a shortened form in Appendix 6.5.
6.2.4 ‘Accent shift’

Perhaps the most interesting examples of a mismatch between measure and morpheme peaks are those with monosyllabic suffixes such as -na (nominaliser).

According to a count from Hawaiian dictionary and Hawaiian grammar, there are only eight suffixes that consist of one light syllable. As mentioned above, these suffixes have the effect of ‘shifting the accent’ one syllable to the right — or, more accurately, shifting the material under the prosodic umbrella. An example is moé-na ‘mat’, from móe ‘sleep’. This type of suffixation is the only morphological process that results in an accentual mismatch between the base form and the derived form. Prefixes and markers that precede a content morpheme never change its accent pattern.

Both Lorrin Andrews (1854) and W.D.W. Alexander (1864) described the behaviour of one such suffix, -la (emphatic), on the word it was attached to. As examples, Andrews (1854:21) compared kiú ‘spy’ with kiú-la. Alexander (1864:22) wrote:

... La unites with the ‘directives’ so as to form one word with them in pronunciation, and after aku, iho, and a’e it shifts the accent to the last syllable, as ihóla, akúla, a’ela.

Other such forms are -a and -na (PASS/IMP; the latter said to be rare), and four transitivisers: -hi, -ki, -'i, and -i.

What is the frequency of the -(C)V suffixes? That of -na seems to be high: Elbert and Pukui (1979:80) considered it the ‘most common’ of the nominalising suffixes and noted that it could appear with ‘many bases’.

These suffixes change the accent of the base form only when the preceding syllables are short (e.g. áku; akúla). If the base form ends in a long vowel or a diphthong, it retains its accent (e.g. papā ‘sound’; papā-na ‘sounding’ (Elbert and Pukui 1979:81); ‘ái ‘eat’, ‘dí-na ‘meal’).

7 Summary

Ha‘ina ‘ia mai (ana) ka puana. ‘Let the summary refrain be told.’

In this study, I have examined Hawaiian accent from a functional point of view. This approach is the opposite of those that posit rules relying on morphology, arbitrary word divisions, or syllable counts that sometimes move from right to left, and at other times, from left to right. I suggest that accent patterns are instead a product of combining two main types of measures — even and uneven — in any order. With this approach, there are no exceptional forms, and there are no stray syllables — a concept that is necessary to make a word such as wahi‘e ‘woman’ fit into a trochaic analysis.

As for the function of accent, my hypothesis was that it serves the hearer as a guide to morphology. To test the hypothesis, I examined the extent to which measures and morphemes in connected speech correspond. The data were phrases and sentences, not the customary lists of words.

When I compared the borders of the units, only 42% of measure boundaries matched those of morphemes. (Following a suggestion by one reader that originally the demarcative function pertained to words, not to morphemes, I found that only 31% of the measures in the data corresponded to words.) However, when I compared the peaks, that is, the

---

I have excluded two possible forms, -u and -‘u. These suffixes are not added to content morphemes, but figure only in forming certain pronouns.
accented syllables in the respective units, there was a much higher match: 84.5%. The implication is that in most cases, the hearer can use measure peaks as guides to morpheme peaks — not a surprising conclusion, since the peaks are easier to hear than the borders.

The study also describes the complex relationship between measure and morpheme demonstrated by the behaviour of light function morphemes with respect to the content morpheme. Those preceding an even measure are incorporated into the measure, producing an uneven measure, as in

\[ \text{[mānu]} \quad \text{‘bird’} \quad \text{[ka mānu]} \quad \text{‘the bird’} \]

but maintaining the accent on the peak syllable of the base morpheme. Those preceding an uneven measure produce a realignment of syllables into two measures,

\[ \text{[wahine]} \quad \text{‘woman’} \quad \text{[kà wa][híne]} \quad \text{‘the woman’} \]

but still keeping the accent on the peak syllable of the base form. The only change in the location of the peak is that produced by light suffixes. For example,

\[ \text{[mōe]} \quad \text{‘sleep’} \quad \text{[moé-na]} \quad \text{‘mat’.} \]

Finally, I have shown that describing Hawaiian accent in context is much more complicated than it is if one deals only with the written headwords in a dictionary. Using words in citation form as data gives a skewed picture, since it cannot treat the realignments of measures that occur when a content form is surrounded by grammatical markers. However, this study is only preliminary and confined to fairly formal speech. What is needed now is a study of natural, less formal speech. 18

Appendix

In this notation the following conventions apply.

1. Measures are enclosed in brackets 19
2. \( M = \) content morpheme
3. \( m = \) function morpheme (light syllable)
4. \( m = \) function morpheme (heavy syllable or disyllable)
5. \( M = \) part of a content morpheme that serves as a measure or part of a measure. Thus, \( M + M = M. \)
6. Hyphens separate morphemes (insofar as we can identify them) within a word.
7. Two spaces separate words. 20
8. The last line contains a translation.

---

18 As a model, I suggest Apolonia Tamata’s analysis of casual speech in Fijian (1994).
19 Other works have separated measures with periods or raised periods. But here, because the notation is complicated, brackets make the measures easier to see.
20 Word division in Hawaiian (as in many other languages) is often orthographically inconsistent, especially when content morphemes are combined in compounds. For example, compare maka ‘upena ‘net mesh’ with maka‘upena ‘midriff’. In the present work, a word is defined as a form that occurs as a headword in Pukui and Elbert 1986.
9. A tentative notation for reduplication interprets a fully reduplicated form as MM, and a partially reduplicated form as either mM or mM, depending on the length of the first syllable. This is not entirely satisfactory; one could also interpret the partially reduplicated form as M.

10. The lengthening of the antepenultimate vowel in a small set of words to denote plural is a notational problem, since this feature is more prosodic than linear (see 6.3 for an example kaikamāhine ‘girls’).

1 Akua

1.1 [He ki’i] [lā][’au] [kē-][ia].
   m M M M m m
   ‘This is a wooden image.’

1.2 [He ki’i] [akua] [kē-][ia].
   m M M m m
   ‘This is a god image.’

1.3 [’O ke-ia] [ki’i] [he hō-][’ai][lōna] [o ke a][kua] [’o Kū].
   m m m M m m M m m M m m M
   ‘This image is a symbol of the god Kū.’

1.4 [I ka] [wā] [kahiko], [’o ke] [ki’i] [he hō][’ai][lōna] [o ka] [mana] [o nā] [kua].
   m m M m m M m m M m m M
   ‘In ancient times, the image was a symbol of the power of the gods.’

2 Ali‘i

2.1 [He po’e] [ali’i] [ke-ia].
   m M M m m
   ‘These are chiefly persons.’

2.2 [’A-‘ahu] [ke a][li’i] [kāne] [i ka] [’ahu-][’ula].
   m M m M M m m M M
   ‘The male chief is clad in a cloak.’

2.3 [Aia] [ka mahi][ole] [ma luna] [o ko-na] [po’o].
   m m M M m M m m M
   ‘There’s a feathered helmet on his head.’

2.4 [Ua] [hume] [’o ia] [i ka] [malo].
   m M m m m m M
   ‘He’s wearing a loincloth.’

21 Kēia shortened to keia here, and the sequence ei resyllabified, resulting in the diphthong ei.
22 Short for akua. The first vowel coalesced with the previous ā, perhaps to avoid an extra-long vowel.
23 Kēia is shortened to keia, but here, the sequence ei is not resyllabified.
24 Full reduplication could be symbolised by MM, but partial reduplication presents a problem. I have chosen to represent it by m.
25 Although locative nouns are a closed set, they are very much like content words, in that they can serve as the nucleus of a grammatical phrase.
26 O and i resyllabify across morpheme boundaries here, see the discussion in §6.2.2.
Measures and morphemes: a functional approach to Hawaiian accent 417

2.5 [Aia] [ka pe][‘ahi] [i ka] [lima] [‘ā][kau] [o ke a][li‘i] [wahine].
   m m M M m m M M m M M M M
   ‘There’s a fan in the right hand of the chiefess.’

2.6 [Ua] [lei] [‘o ia] [ka lei] [niho] [palaoa].
   M m m m M M M M M M
   ‘She’s wearing (as a lei) the whale’s tooth pendant.’

2.7 [Ua] [kau] [ka lei] [hulu] [ma luna] [o ke] [po‘o].
   M m M M m M m M M M
   ‘A feather lei is placed on her head.’

3 Kahuna

3.1 [He ka][huna] [kē-][ia].
   M M m m
   ‘This is a priestly expert.’

3.2 [He mana] [ika-ika] ko-na].
   M M M m m
   ‘He has strong spiritual power.’

3.3 [He nui] [na ‘ano] [kā][huna] [i ka] [wā] [kahiko].
   M M m M M M M m M M
   ‘There were many kinds of priestly experts in olden times.’

4 Maka‘āinana

4.1 [He po‘e] [maka-][‘āi][na-na] [kē-][ia].
   M M M M m m m
   ‘This is a commoner.’

4.2 [He ka][naka] [mahi-][‘ai] [paha] [ke kāne].
   M M M M m m M M
   ‘Perhaps the man is a farmer.’

4.3 [Aia] [ka ‘ō][‘ō] [i ko-na] [lima] [‘ā][kau].
   m m M M m m m M M M M
   ‘There’s a digging stick in his right hand.’

4.4 [Ua] [hume] [‘o ia] [i ka] [malo].
   M m m m m m M
   ‘He’s clad in a malo.’

4.5 [Noho] [ka wa][hine] [i lala].
   M M m M M m
   ‘The woman is sitting down.’

27 Note the resyllabification in ikaika, which shortens the form to one measure.
28 The underlying form is [nā] [‘ano]. As mentioned in Elbert and Pukui 1979:14, a glottal stop can shorten a preceding vowel.
4.6 [Ua] [kōma] [ʻo ia] [i ka] [pā] [ʻū].

m M m m m M M

‘She’s wearing a sarong.’

4.7 [Ua] [lō] [ʻīhi] [ko-na] [lau-] [oho].

m M m m M M

‘Her hair is long.’

4.8 [He wa] [hi] [uʻi] [ʻo ia].

m M M m m M

‘She’s a beautiful woman.’

5 Kauā

5.1 [He poʻe] [kau] [ā] [kē-] [ia].

m M M m m

‘These people are outcasts.’

5.2 [Ua] [pio] [paha] [lā-ua] [i ke] [kaua].

m M m m m m m M

‘Perhaps they were captured in a war.’

5.3 [Ua] [kā][kau] [ʻia] [lā-ua] [i ka] [hikoni] [ma ka] [hele-] [hele-na].

m M M m m m m M m M m M m

‘They were tattooed with brands on their face.’

5.4 [He hō-][ʻai][lōna] [kē-][lā] [o ko] [lā-ua] [kū-][lana].

m M M M m m m m m m M m

‘That was a symbol of their status.’

6 Inoa

6.1 [ʻO nā] [inoa] [ke kumu-][hana] [o ke][-ia] [kiʻi].

m m M m M m m m M m M

‘Names are the subject of this picture.’

6.2 [ʻO Ke][ko] [ka i][noa] [o ke] [keiki] [kāne].

m M m m M M m m M M m

‘Kekoa is the name of the boy.’

6.3 [ʻO Kuʻu-][le][lā-ua] [ʻo Moki][hana] [nā] [inoa] [o nā] [kaika-][mā][hine].

m m M m m m M M M m m M m M m M m M M

‘Kuʻulei and Mokihana are the names of the girls.’

29 Note the resyllabification in lāua.
30 The underlying form is [o kē][ia].
31 Note the resyllabification in ka inoa.
32 This sequence of measures (a heavy syllable followed by an uneven measure) I have labelled ‘indeterminate’, since it could be divided either as [nā] [inoa] or [nāi] [noa]. In the measure marking in the Hawaiian dictionary, Elbert has chosen the former division for all such forms.
33 This is an example, noted above, of a non-linear morpheme (PL), difficult to handle in this system of notation.
6.4 ['O wai] [kou]\textsuperscript{35} [inoa]?
\begin{itemize}
\item m m m M
\end{itemize}
‘What’s your name?’

6.5 ['O wai] [ka i][inoa] [o ka i-][ilio]\textsuperscript{36}
\begin{itemize}
\item m m M M m m M M
\end{itemize}
‘What’s the name of the dog?’

Glossary of words in the appendix

| ‘a’ahu    | ‘clad’       | komo  | ‘wear’     |
| ‘ahu’ula  | ‘feather cloak’ | kona  | ‘his, hers, its’ |
| ‘ākau     | ‘right-hand’  | kou   | ‘your’     |
| ‘ano      | ‘type’       | Kū    | name of god  |
| ‘ia       | passive marker | Ku’ulei | person’s name |
| ‘o        | topic, proper marker | kālana | ‘status’   |
| ‘ō’ō      | ‘digging stick’ | kumuhana | ‘subject’ |
| aia       | ‘there is’   | lā’a’u | ‘wooden’   |
| akua      | ‘god’        | lalo  | ‘down’     |
| ali’i      | ‘chief’     | lāua  | ‘they two’ |
| he        | indefinite, existential marker | lauoho | ‘hair’ |
| helehelena | ‘face’      | lei   | ‘garland’ |
| hikoni    | ‘brand’     | lima  | ‘hand’     |
| hō’aïona  | ‘symbol’    | lō’ihi | ‘long’    |
| hulu      | ‘feather’   | luna  | ‘above’    |
| hume      | ‘wear’      | ma    | ‘at’       |
| i         | object marker, ‘to, for, in’ | mahi’ai | ‘farmer’ |
| ia        | ‘he, she, it’ | mahiole | ‘helmet’ |
| ikaika    | ‘strong’    | maka’aïinana | ‘commoner’ |
| i‘ilio    | ‘dog’       | malo  | ‘commoner’ |
| inoa      | ‘name’      | mana  | ‘power’    |
| ka        | definite article | Mokihana | person’s name |
| kahiko    | ‘olden’     | nā    | plural marker |
| kahuna    | ‘expert’    | niho  | ‘tooth’    |
| kāhuna    | ‘experts’   | noho  | ‘sit’      |
| kaikamāhine | ‘girls’  | nui   | ‘many, large’ |
| kākau     | ‘tattoo’    | o     | possessive |
| kanaka    | ‘person’    | pā’ū  | ‘skirt’    |
| kāne      | ‘male’      | paha  | ‘perhaps’  |
| kau       | ‘placed’    | palaoa | ‘whale’s tooth’ |
| kaua      | ‘war’       | pe’a hi | ‘fan’     |
| kauā      | ‘slave’     | pio   | ‘captured’ |
| ke        | definite article | po’e  | ‘people’ |
| kēia      | ‘here, this’ | po’o  | ‘head’     |
| keiki     | ‘child’     | u‘i   | ‘beautiful’ |
| Kekoa     | person’s name | ua   | perfect aspect marker |
| kēlā      | ‘there, that’ | wā    | ‘time’     |
| ki‘i       | ‘image’     | wahine | ‘woman’   |
| ko        | possessive | wai   | ‘water’    |

\textsuperscript{35} It might be possible to analyse kou as consisting of two morphemes.

\textsuperscript{36} ‘ilio is shortened here to ‘ilio.
References


The optional ergative in Kâte

EDGAR SUTER

1 Introduction

Kâte has a case enclitic to mark the subjects of transitive verbs. This marker, however, does not automatically attach to all transitive subjects. In spontaneous discourse transitive subjects are sometimes found to be marked with this optional ergative and sometimes they are unmarked, as in (1) and (2).

   1s-ERG feat like.this do-HAB.PRS.1s
   ‘I usually do this kind of thing.’

      SEQ.DS 3s feat like.that do-RPST.3s
      ‘Then he did that kind of thing.’

(2) a. Eme [mamac-nane-zi]₁ mu-wec: ...
      SEQ.DS father-1s.POSS-ERG say-RPST.3s
      ‘Then my father said: ...’

   b. Eme biac [no-re mamac]₁ muwec: ...
      SEQ.DS soon 1s-GEN father say-RPST.3s
      ‘Then my father said: ...’

(1a) and (1b) were uttered by the same speaker in the course of a discussion. The expression mâsi e, lit. ‘perform a miracle’, is a faded hyperbole that means hardly more than ‘do something’ in the colloquial register employed in this conversation. Both sentences in (1) have a personal pronoun as their subject, but whereas the pronoun in (1a) is marked with the ergative, that in (1b) is not. Similarly, of the two synonymous sentences

---

1 This paper was drafted in 1998 when I was one of Andy Pawley’s PhD students in Canberra. I remember with fondness the many times Andy and Medina invited me (and other guests) to their home for dinner during the first half of my term when my wife and I were living apart. It is a pleasure to devote this fragment of my thesis to a mentor who supported me in difficult times. I am grateful to Cindi Farr and Bill McGregor for valuable comments. I am indebted to my main informants Mupenarec Fârepe and John Kembarang who discussed the example sentences included in this paper with me. The responsibility for all remaining inaccuracies is of course mine.
in (2) the subject of the first is marked with the ergative whereas the second is unmarked. The two pairs of sentences in (1) and (2) are as close as we get to finding minimal pairs differing only in the presence or absence of the ergative in the corpus of discourse underlying this study. When confronted with sentences such as these, informants agree that the ergative marker may just as well be present or absent without impinging on the acceptability of the sentences. The reason why the ergative should or should not be used in discourse is not open to introspection.

Kâte is a typical Trans New Guinea language in that the major method of indicating grammatical relations is affixation to the verb. There are two sets of object suffixes, one registering person and number of the direct object, the other that of indirect objects and beneficiaries. In an outer layer we find the subject suffixes, which are fused with the morphemes for tense and mood. All three of these suffix sets register the person and number of human referents; inanimate referents trigger the form of the third singular. The verb morphology of Kâte therefore follows an accusative pattern of marking the grammatical relations. Noun phrases with a nominal head being cross-referenced by these suffixes are generally not marked for person or number. In the case of the object relations, there is no case enclitic that can mark a noun phrase. Direct and indirect object NPs are always unmarked. Only the subject NP can be marked by a phrasal enclitic (-zi) that is otherwise used to signal the instrumental case.

Saying that this case marker optionally occurs on the subject NP of transitive clauses is of course rather imprecise. This paper is devoted to the task of delineating the conditions of use of this marker in some detail. This requires several steps. Section 2 presents the basic morphological and syntactic facts. In §3 we turn to discourse analysis and look at the distribution of ergative marking in monologue and dialogue. Section 4 expands on the result of §3 by studying the interaction of the ergative with focus particles and interrogative pronouns. Section 5, finally, summarises the major findings of the investigation and places them in the context of the optional ergatives of other Trans New Guinea languages.

2 Morphology and syntax of ergative marking

In his pioneering grammar of Kâte, Pilhofer (1933) gave an excellent account of the form and the main uses of the optional ergative (which he called Nominativ agentis). In this section I will start with a summary of his findings, illustrating them with my own data.

Before we turn to syntax, a look at the morphology of the ergative is in order. The ergative enclitic -zi occurs on nouns, but cannot be attached to personal pronouns. Instead, there are special ergative forms of the personal pronouns. However, such ergative forms only exist for the singular pronouns, the dual and plural pronouns lack them and have only a single, unmarked form (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>SG</th>
<th>DU</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>Unmarked no</td>
<td>Ergative noni</td>
<td>Exclusive: nâhe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Inclusive: nâhâc</td>
</tr>
<tr>
<td>2nd</td>
<td>go</td>
<td>goki</td>
<td>ŋohe</td>
</tr>
<tr>
<td>3rd</td>
<td>e</td>
<td>eki</td>
<td>jahe</td>
</tr>
</tbody>
</table>
The singular and the non-singular pronouns behave differently in contexts in which ergative marking is preferred. Example (3) shows the same sentence with a pronominal subject in the plural (3a) and in the singular (3b).

(3) a. \[Neŋgoc \text{ e-kicne } yoge(*-zi) \_A \text{ jāmbāŋ} \{gie\}_O \text{ ro-niŋ.}\]
mother become-PART 2p(-ERG) truly work take-NHORT.2p
‘You who are mothers, you should get the job.’

b. \[Neŋgoc \text{ e-kicne } go-ki / (?go) \_A \text{ jāmbāŋ} \{gie\}_O\]
mother become-PART 2s-ERG (2s) truly work
\text{ro-c.}
take-NHORT.2s
‘You who are a mother, you should get the job.’

(3a) was found in discourse with a subject NP that ends in the unmarked second plural pronoun \text{yoge}. If we transpose this sentence into the singular number (3b), we see that there is a strong preference for the subject of this sentence to host the ergative marker (cf. §4). Despite the clear preference for the ergative in this structure, the plural pronoun in (3a) cannot be so marked. It is a morphological fact that the dual and plural personal pronouns have no ergative form. The same holds for the emphatic pronouns (\text{no nahac} ‘I myself’, etc.; cf. Pilhofer (1933:53)), but in the case of the emphatic pronouns even the singular pronouns have no ergative form, as illustrated in (4).

(4) a. \[Hectar \text{ piicne mac jāmbāŋ } zi \text{ orāṅke-kac } i]_O\]
plot little DIM truly this go.around-PRS.3 that
\[\text{misiŋ e } \text{ jahac(*-zi)} \_A \text{ ro-wec.}\]
mission 3s 3s.EMPH(ERG) take-RPST.3s
‘The mission itself claimed the small area of land around here.’

b. \[Hectar \text{ piicne mac jāmbāŋ } zi \text{ orāṅke-kac } i]_O\]
plot little DIM truly this go.around-PRS.3 that
\[\text{misiŋ-zi} / (*\text{misiŋ}) \_A \text{ ro-wec.}\]
mission-ERG (mission) take-RPST.3s
‘The mission claimed the small area of land around here.’

Kâte is an SOV language with rigid word order. The order of the argument NPs in a transitive clause is relevant for their correct syntactic interpretation. Ambiguity might arise in transitive clauses with a verb (such as ‘hit’ or ‘see’) that allows both arguments to be human. If one presents informants with such clauses in which both NPs are headed by unmarked nouns, they use word order to interpret them: the first NP is assigned the subject role, the following NP the direct object role (see (5a)). SOV is the basic word order and in discourse the great majority of transitive clauses with two argument NPs have this order. The only other order of constituents that is possible without breaching the integrity of the clause is OSV (see (5b)). To make informants perceive an OSV clause in a potentially ambiguous situation as just described, it is necessary to mark the second consecutive NP with the ergative. OSV is quite literally a marked word order in that the subject in such clauses must be marked with the ergative, as was already noticed by Pilhofer (1933:125).
(5) a. SOV: SBJ {-ERG DO V PRED (ERG-marking optional)}

b. OSV: DO SBJ-ERG V PRED (nearly always ERG-marked)

In the corpus of discourse studied for this paper\(^2\) there were three exceptions to the rule represented in (5b). Given that the number of OSV clauses with an unmarked subject is so small I refrain from speculating about the reasons for these exceptions. It is worth noting, however, that informants, when confronted with such clauses, consider them problematic and have suggested improving them by adding the ergative marker. Given this response and the rarity of the exceptions, I think we may consider the necessity to mark the subject in OSV clauses with the ergative a syntactic rule, while at the same time admitting that the rule leaks.

Although the optional ergative of Kâte is found on transitive subjects in the vast majority of its occurrences in discourse, this is not the only syntactic environment in which it may occur. Sporadically it is also found on the subjects of intransitive verbs.\(^3\) Pilhofer (1933:104) noted that in such clauses the ergatively marked intransitive subject is often in contrast with another referent. Indeed it is not difficult to find clauses that exemplify this observation.

(6) No-re kec qâreŋke-râ ine Sopâc i-rec mu-mbiy:
  1s-DEST thus write-SEQ SS but Sopâc that-DEST say-RPST.3p
  zi gie afecne, i-rec [go-ki]s ju-c. No-re
  this work few that-DEST 2s-ERG be-NHORT.2s 1s-DEST
  mu-mbiy: Go ine wemo ra-naymu, â Buka ra-naymu.
  say-RPST.3p 2s but what go-FUT.1p um Buka go-FUT.1p
  ‘As [the Australian officers] took notes about me, they said concerning
  the man from Wândokai: Here there is little work, so you stay. Concerning
  myself they said: You will go to — what’s it called — to Buka with us.’

Example (6) is taken from an account of the Second World War on the Huon Peninsula. In the preceding context the narrator has just described how he worked together with another local man on an Australian base. Then, in (6), he relates how the two were forced to part from one another. In the command goki juc ‘you stay!’ the referent of the pronoun, the man from Wândokai, is clearly opposed to the narrator.

In some other Trans New Guinea languages with optional ergatives, intransitive verbs have been described as falling into two groups, one consisting of verbs that can only take subjects in the absolutive case, the other consisting of verbs that variously take absolutive or ergative subjects (e.g. Tauya (see MacDonald 1990:317) and Fore (Scott 1986:172)). In Folopa (Anderson and Wade 1988:6) the difference between these two groups of intransitive verbs seems to depend on the agentivity potential of the subject. Only verbs

\(^2\) This corpus encompasses about 50,000 words, i.e. it is considerably larger than the part of it reported in §3.

\(^3\) The occurrence of \(-zi\) on the subjects of intransitive verbs could be taken as an indication that this is really a nominative rather than an ergative case. This is, in fact, the position adopted by Donohue and Donohue (1997) in their discussion of a similar marker in the Fore language. I will address this terminological question, which is not germane to the descriptive concerns of this paper, in §5.
whose subject has some control over the action can co-occur with the ergative. This is in line with the contrast seen in transitive verbs where use of the optional ergative signals that the agent instigates and controls the action (Anderson and Wade 1988:9). In Kâte no such agentivity effects have been observed either in transitive or in intransitive clauses. Ergative marking does not appear to be limited to a particular subclass of intransitive verbs (it is, however, sufficiently rare to have been observed only with a small number of such verbs). In particular, the ergative also occasionally occurs on intransitive subjects that can by no stretch of the imagination be considered agentive.

(7)  
\[ [Australia-zi]_S \text{ mâreŋ} juwickec irec woraj fuŋ-ko \]
and Australian-ERG land down.there from.there mango base-LOC

\[ kec irec hâmo-mbiŋ. [Ŋic-zi Amerika â Australia]_S \]
lo! from.there die-RPST.3p man-ERG American and Australian

\[ hâmo-mbiŋ. \]
die-RPST.3p

‘The Australians died over there where you can see the mango tree.
American and Australian men died.’

Example (7) is taken from the same wartime narrative as (6). The narrator has just related how the Allied forces trapped the Japanese. Then, in a somewhat surprising turn of perspective, he mentions the casualties of the Allied forces. Thus the intransitive subjects in (7) are in a sense also contrastive, as we would rather have expected to hear of the death of the Japanese. The second of the two sentences in (7) is an amplification of the first in which the narrator forgot to mention the Americans. In both these sentences the predicate is \( hâmo \) ‘die’, a verb whose subject certainly has no control over the event.

Pilhofer (1933:104f.) mentions a related function of the ergative in transitive clauses. Adding the ergative marker can lay emphasis on the subject. The absence of the ergative, on the other hand, may signal that the subject is not prominent. Evidently, Pilhofer is talking about the information structure of clauses here. I will take up his hint and explore the interaction of the ergative with information structure in §3 and §4. It must be said here, however, that the focussing function of the ergative is much subtler in transitive than in intransitive clauses. Most instances of the ergative in transitive clauses are like examples (1) and (2) in which the focussing function is not immediately obvious. Examples in which a transitive subject is in contrast with another contextually given referent are not numerous. In intransitive clauses, on the other hand, the use of the ergative frequently signals such a contrast, as in (6) and (7).

3 The distribution in discourse

Pilhofer, like a number of other grammarians describing languages with an optional ergative (e.g. Bromley 1981; Haiman 1980), sets great store by the disambiguating function of the ergative case marker. He gives examples of ambitransitive verbs where the ergative can distinguish between a transitive and an intransitive reading, and examples of transitive clauses with a single overt NP where the ergative can distinguish between the subject and an object role (Pilhofer 1933:103f.). All these examples, however, are clearly elicited and reflect the conscious manipulation of the ergative by informants when they are made aware of a problem (cf. example (8) below). I do not doubt that the ergative can have a disambiguating effect in spontaneous discourse, too, but I very much doubt if this is its
raison d’être. Speakers of a language with an optional ergative certainly do not monitor their speech and insert an ergative marker whenever a syntactically ambiguous structure threatens to surface and otherwise omit the marker. If this were so, there would never be any need to use the ergative in SOV clauses in Kâte, since here word order already signals the grammatical relations of the arguments. For all we know, speakers are generally unaware of syntactic ambiguity and do not worry about it. Nevertheless, it is an interesting question whether there is any evidence in spontaneous discourse of a differential treatment of transitive clauses with a single NP and such with two NPs. In Kâte, the need to mark the grammatical relation of an NP is greatest in transitive clauses with a single overt NP, since here word order does not help in the assignment of the intended role.

I made a quantitative discourse analysis of the use of the ergative in texts belonging to two different genres, one monological, the other dialogical. The dialogue is a discussion of 40 minutes duration about local politics. In this discussion four young men participated, although the contribution of one of the participants was limited to occasional comments. Monologue is represented by two autobiographical accounts, each taking up a little more than 20 minutes recording time. The speakers were two middle aged men. Table 2 gives a summary of the transitive clauses that occurred in the texts of these two genres.

Table 2: Representation of arguments in transitive clauses

<table>
<thead>
<tr>
<th>SBJ</th>
<th>DO</th>
<th>Discussion</th>
<th>Autobiography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>NP</td>
<td>218</td>
<td>364</td>
</tr>
<tr>
<td>Ø</td>
<td>Ø</td>
<td>82</td>
<td>201</td>
</tr>
<tr>
<td>NP</td>
<td>NP</td>
<td>109</td>
<td>99</td>
</tr>
<tr>
<td>NP</td>
<td>Ø</td>
<td>48</td>
<td>33</td>
</tr>
<tr>
<td>Total TR clauses</td>
<td>457</td>
<td>100%</td>
<td>697</td>
</tr>
</tbody>
</table>

The first thing we notice when we look at Table 2 is the difference in the number of transitive clauses in the discussion and in the autobiographies. Even though the discourse samples of these two genres took up roughly the same time, there are one and a half times as many transitive clauses in the autobiographies as in the discussion. The distribution of overt NPs is also characteristically different. What is the same across both genres is the ranking of single NP clauses. In both conversation and narrative, clauses with an overt direct object NP and a zero subject are far and away the most frequent, coming to roughly half of all transitive clauses. Clauses with only a subject NP, on the other hand, are the least frequent type in both genres. The difference lies in the middle ground covered by double NP and double zero clauses. In autobiographical narrative, clauses with two zero arguments are twice as common as double NP clauses. In conversation, however, the ranking is reversed with double NP clauses being more frequent. This, together with the fact that subject-only clauses are twice as frequent in conversation as in narrative, is responsible for the higher number of overt subject NPs in the discussion, despite the smaller number of transitive clauses. In the discussion a third of all transitive clauses have a subject NP, in the autobiographies only a fifth.

Now we will look at the distribution of the ergative in clauses with one and two NPs. Since we are concerned with the marking of the subject NP, only single NP clauses with an
overt subject NP are of relevance. Table 3 shows the distribution of ergatively marked and unmarked subject NPs in clauses with and without a direct object NP.4

**Table 3:** Ergative marking and the representation of the direct object

<table>
<thead>
<tr>
<th>SBJ</th>
<th>DO</th>
<th>Discussion</th>
<th>Autobiography</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>Ø</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>NP-ERG</td>
<td>Ø</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>NP</td>
<td>NP</td>
<td>39</td>
<td>42</td>
</tr>
<tr>
<td>NP-ERG</td>
<td>NP</td>
<td>21</td>
<td>27</td>
</tr>
<tr>
<td>Total ERG-markable TR clauses</td>
<td>90</td>
<td>100%</td>
<td>94</td>
</tr>
</tbody>
</table>

When we turn our attention to the different types of clauses in Table 3, we notice that ergatively marked and unmarked subject NPs pattern differently in single NP and in double NP clauses. In clauses with an overt direct object NP, unmarked subjects are considerably more frequent than subjects marked with the ergative both in monologue and in dialogue. In clauses with an ellipsed direct object NP, by contrast, ergatively marked subjects are more frequent than unmarked subjects, albeit only marginally so in dialogue. Thus we do find evidence in spontaneous discourse that ergative marking is preferred in single NP clauses whereas it is eschewed in double NP clauses. However, this is not the whole story. Before we proceed to an interpretation of this finding we need to examine some more figures.

It is worth looking at the total number of ergatively marked NPs in Table 3. Ergative marking is somewhat more frequent in the autobiographies with 48% of all transitive subject NPs being marked, as opposed to 41% in the discussion. These figures refer to subject NPs that can be marked by the ergative. If we go back to Table 2 and also include the subject NPs that cannot be ergatively marked, the percentage drops. Only 24% of all overt transitive subject NPs in the discussion and 34% in the autobiographies are in the ergative case. Thus, non-marking of the transitive subject is more frequent than marking in Kåte.

**Table 4:** Ergative marking on nominal and pronominal subjects

<table>
<thead>
<tr>
<th>TR SBJ</th>
<th>Discussion</th>
<th>Autobiography</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>21</td>
<td>27%</td>
</tr>
<tr>
<td>N-ERG</td>
<td>24</td>
<td>31%</td>
</tr>
<tr>
<td>PRON.SG</td>
<td>20</td>
<td>26%</td>
</tr>
<tr>
<td>PRON.SG-ERG</td>
<td>13</td>
<td>17%</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>100%</td>
</tr>
</tbody>
</table>

4 A look at the total numbers of clauses at the bottom of Table 3 shows that fewer clauses were considered than appear in Table 2 as having a subject NP. The reason for this is the fact, mentioned in §2, that there are pronouns which do not have an ergative form. In Table 3 only subject NPs were included which either were in the ergative or could potentially be marked with the ergative.
Next, we want to compare nominal and pronominal subjects. The fact that some personal pronouns, that is, the dual and plural ones, cannot be marked with the ergative case suggests that personal pronouns in general may eschew the ergative. Table 4 breaks down the transitive subjects into those with a nominal and those with a pronominal head.\(^5\) Again only ergative-markable NPs were considered, i.e. only singular pronouns. The figures show that non-marking of pronominal subjects is indeed preferred both in dialogue and in monologue. Nominal subjects, on the other hand, are more often marked than unmarked. Both these tendencies are very pronounced in monologue, but less clear in dialogue.

Is there any common factor that might be responsible for the propensity of subjects in clauses with a zero direct object (Table 3) and subjects with a nominal head (Table 4) to take ergative marking? I believe there is. Remember that subjects in OSV clauses are nearly always marked with the ergative case. Now a frequent — although not the only — reason for choosing OSV word order is the topicalisation of the direct object. There is a trade-off between thematic and rhematic constituents in the same clause: if the direct object is topicalised, the subject will become more prominent. I suggest that the same mechanism is at work in clauses with an overt subject and a zero direct object. The reason for eliding the direct object is that it is thematic, which goes hand in hand with the subject being more rhematic. The difference between nominal and pronominal subjects fits in well with such an account. Personal pronouns are inherently more thematic than nouns. Thus in all three structures concerned we find that heightened rhematicity of the subject correlates with more frequent ergative marking. Conversely, when the subject is less rhematic — in SOV clauses, in clauses with an overt direct object NP, and in clauses with a pronominal subject – ergative marking is less frequent.

### Table 5: Both criteria combined (cf. Tables 3 and 4)

<table>
<thead>
<tr>
<th>SBJ</th>
<th>DO</th>
<th>Discussion</th>
<th>Autobiography</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ERG : all</td>
<td>ERG</td>
</tr>
<tr>
<td>N</td>
<td>Ø</td>
<td>11 : 17</td>
<td>65%</td>
</tr>
<tr>
<td>N</td>
<td>NP</td>
<td>13 : 28</td>
<td>46%</td>
</tr>
<tr>
<td>PRON.SG</td>
<td>Ø</td>
<td>5 : 11</td>
<td>45%</td>
</tr>
<tr>
<td>PRON.SG</td>
<td>NP</td>
<td>8 : 22</td>
<td>36%</td>
</tr>
</tbody>
</table>

Table 5 shows the ranking of the different clause types if both the criteria given in Tables 3 and 4 are combined. The clause type with the highest propensity to be marked ergatively has a nominal subject and a zero direct object (11 out of 17 such structures, or 65%, were marked with the ergative in discussion). At the other end of the scale are clauses with a pronominal subject and an overt direct object. Note that the ranking of the clause types combining these two criteria is the same in dialogue and in monologue. Yet there is a marked difference between the percentage figures of the two discourse genres. In monologue the two clause types at opposite ends of the scale are maximally divergent, showing 93% and 14% of ergative marking, respectively. In dialogue, on the other hand,

\(^5\) The total number of transitive subjects considered is again a bit smaller than in the preceding table. This is due to the occasional occurrence of NPs exhibiting pronominal copy, such as *njic e ‘man he*. Such NPs were counted neither as nominal nor as pronominal and hence were excluded from Table 4.
the range of variation is compressed, reaching from 65% down to 36%. I have no explanation for this striking difference in the distribution of the ergative between monologue and dialogue. However, it is not implausible to assume that the rhematicity of subjects is distributed differently across clause types in the two discourse genres. At any rate, the difference in the distribution of the ergative between monologue and dialogue strongly suggests that the use of the ergative in Kâte is governed by pragmatic factors.

Now let’s return to the finding that transitive clauses with a single overt (subject) NP are more likely to have an ergatively marked subject than clauses with two overt NPs. Pilhofer and like-minded grammarians might see in this a confirmation of their view that the ergative serves a disambiguating function. This view seems to be further corroborated when we consider the reaction of native speakers to sentences like those in (8).

After what we have seen about the optionality of the ergative case in discourse, we might expect sentence (8a) to have two possible interpretations, the one given and the same one as (8b). However, informants are reluctant to accept ‘The man saw her.’ as a possible interpretation of (8a). When I asked if this was possible, I was invariably told that (8b) should be used to express this meaning. Thus native speakers have strong intuitions about the correct interpretation of transitive clauses with a single overt argument when the semantics of the verb and the argument allow two different role assignments: if the argument is marked with the ergative, it is the subject; if it is unmarked, it is the direct object. What could be the reason for this intuition? If we go back to Table 5 we see that, in narrative discourse, clauses of the type of (8) almost categorically require ergative marking if the single noun is to be interpreted as the subject (14 out of 15 such structures had ergative marking). I suspect that informants treat isolated sentences such as those in (8) as mini-narratives when they are asked to think about them. In other words, they fall back on the usage that is appropriate for narrative discourse when they say that sentences such as (8b) require ergative marking. It is therefore not necessary to invoke a desire to avoid ambiguity, although this may be an additional factor that comes into play in an elicitation situation of this sort.

4 Interaction with focus particles and interrogative pronouns

In the course of the analysis of discourse in the preceding section we arrived at the conclusion that it is the rhematicity of the subject that triggers the use of the ergative case. In this section I want to put this finding to the test. We will see how transitive subjects behave when they occur in a highly rhematic context. Such a context is provided by focus particles. Kâte has five focus particles: jaha ‘EMPH’, sawa ‘only’, wâc ‘too’, hâcne ‘indeed’ and jàmbâŋ ‘truly’. All of these particles immediately follow the constituent they focus. We see examples of them in (9) through (11), all of them taken from spontaneous discourse.

(9) Wofuŋ-zi hâcne gamaŋ râe-wec.
    Lord-ERG indeed authorities put-RPST.3s
    ‘It was the Lord who put the authorities in place.’
(10) **No-ni sawa mu-pe mu-murâ biayne mi e-ocmu.**
1s-ERG only say-SEQ.1s.DS if good not become-FUT.3
‘If I tell it alone, it won’t be good.’

(11) **E-ki jâmbâŋ wâc mi soŋagke-wec.**
3s-ERG truly too not take.care-RPST.3s
‘He didn’t really take care of it, either.’

In each of (9) through (11) the subject of a transitive verb is focussed by a focus particle or, in the case of (11), by a combination of two focus particles. As we can see, the subject is always in the ergative case. If we try to eliminate the ergative case from the subjects in (9) to (11), this will affect the acceptability of the sentences. Informants decidedly prefer to have a focussed subject marked with the ergative. This can also be seen in (12).

(12) a. **Eme GO-ki jâmbâŋ hâcne jaza-mec!**
SEQ.DS 2s-ERG truly indeed tell.them-NPST.2s
‘So it was really you who told them!’

b. **?Eme GO jâmbâŋ hâcne jaza-mec!**
SEQ.DS 2s truly indeed tell.them-NPST.2s
‘So it was really you who told them!’

The pronoun of the second singular in (12), strongly focussed by two consecutive focus particles, must be marked with the ergative case. Omitting the ergative (12b) results in a sentence that sounds unnatural. To demonstrate that it is indeed the presence of the focus particle that is responsible for the doubtful acceptability of sentences such as (12b), I varied a sentence, including and excluding a focus particle as in (13).

(13) a. **Qeraharuc JU-hapie NO-ni sawa honec-jopa-po.**
Sattelberg be-SIM.3p.DS 1s-ERG only see-3p.DO-RPST.1s
‘Only I saw them when they were in Sattelberg.’

b. **?Qeraharuc JU-hapie NO sawa honec-jopa-po.**
Sattelberg be-SIM.3p.DS 1s only see-3p.DO-RPST.1s
‘Only I saw them when they were in Sattelberg.’

c. **Qeraharuc JU-hapie no-ni Honec-jopa-po.**
Sattelberg be-SIM.3p.DS 1s-ERG see-3p.DO-RPST.1s
‘I saw them when they were in Sattelberg.’

d. **Qeraharuc JU-hapie no Honec-jopa-po.**
Sattelberg be-SIM.3p.DS is see-3p.DO-RPST.1s
‘I saw them when they were in Sattelberg.’

In (13a) and (13b) the subject pronoun is bound by the focus particle sawa ‘only’. Informants respond to these two sentences just as they did to those in (12): only (13a), where the subject is marked with the ergative, is a fully acceptable sentence, whereas (13b) sounds odd. No such difference in acceptability is reported for (13c) and (13d) where the focus particle is missing. Both (13c), with a subject in the ergative, and (13d), with an unmarked subject, are perfectly acceptable sentences.

---

6 In (12) and all following elicited sentences, the prosodically most prominent syllable of the clause (focus accent) is given in small capitals.
These acceptability judgments are entirely consistent with what we find in discourse. Transitive subjects that are followed by a focus particle are nearly always marked with the ergative case. If the direct object is focussed by a particle, on the other hand, the subject of the same clause may either be in the ergative or unmarked. This is summarised in (14).

(14) a. \(SBJ - ERG\) \(FOC\) \(DO\) \(V\) \(PRED\) (nearly always ERG-marked)

b. \(SBJ\) \(-ERG\) \(DO\) \(FOC\) \(V\) \(PRED\) (ERG-marking optional)

Beside OSV clauses we have thus found another environment in which ergative marking is virtually obligatory. The necessity to mark transitive subjects that are bound by a focus particle confirms the conclusion we reached in §3. Again we find that high rhematicity triggers ergative marking.

There is another context of high rhematicity, that is, questions with an interrogative pronoun. Interrogative pronouns are inherently rhematic. In the following we are going to see how interrogative pronouns interact with the ergative and with word order.

(15) a. \([MO] jānē\) \(A\) \([sāec]\) \(O\) muru-mbieŋ?
    who 3p wallaby shoot-NPST.3p
    ‘Who shot the wallaby?’

b. \([sāec]\) \(O\) \([MO] jānē\) \(A\) muru-mbieŋ?
    wallaby who 3p shoot-NPST.3p
    ‘Who shot the wallaby?’

If the interrogative pronoun has subject function and the direct object is a noun, as in (15), both SVO and OSV word order are possible. This holds even if the subject cannot be ergatively marked as in (15), where the interrogative pronoun has been pluralised by appending the personal pronoun of the third plural which cannot host the ergative marker. The situation looks different when we look at transitive clauses that are made up of a subject and a direct object with maximally opposing inherent rhematicity values. In (16) all possible combinations of clauses with hone ‘see’ as predicate and e ‘he, she’ and mo ‘who?’ as arguments, with and without ergative marking, are given. The personal pronoun is inherently thematic, whereas the interrogative pronoun is rhematic.

(16) a. \([E-ki]\) \(A\) \([MO]\) \(hone-wec?\)
    3s-ERG who see:3s.DO-RPST.3s
    ‘Who did she see?’

b. \([E]\) \(A\) \([MO]\) \(hone-wec?\)
    3s who see:3s.DO-RPST.3s
    ‘Who did she see?’

c. \([MO]\) \(\{e-ki\}\) \(hone-wec?\)
    who 3s-ERG see:3s.DO-RPST.3s
    ‘Who did she see?’

d. *\(MO\ e\) hone-wec?
    who 3s see:3s.DO-RPST.3s
Examples (16a) and (16b) have SOV word order and the direct object is the rhematic interrogative pronoun. The situation in these two clauses is comparable to what we found for clauses containing a focus particle, see (14): if the direct object is focussed, ergative marking is optional. Examples (16c) and (16d) show the reversed order of the constituents in the first two examples. (16c) is interpretable but does not sound quite right, whereas (16d) is ill-formed. If pressed, informants give (16d) an SOV interpretation (‘Who saw her?’), but they add that this is not the way to say this. In (16e) and (16f) the interrogative pronoun is marked with the ergative and accordingly is the subject, but only (16e) is accepted as the common way of saying ‘Who saw her?’ whereas (16f) is unusual. This is surprising given that (16f) has SOV word order. Like (16c) and (16d), however, in (16f) the interrogative pronoun is in the first position of the clause where a thematic constituent is expected. This detracts from the acceptability of these utterances. If we look at the examples that are fully acceptable ((16a), (16b) and (16e)), we see that in them the interrogative pronoun is always in second position and the personal pronoun is in first position in the clause. In other words, in these clauses the constituents are ordered according to the sequence theme-rheme. We see the same phenomenon in all clarity in (17).

(17) a. Go mo-zi gaza-jec?
    2s who-ERG tell.thee-NPST.3s
    ‘Who told you?’

b. ?mo-zi go gaza-jec?
    who-ERG 2s tell.thee-NPST.3s
    ‘Who told you?’

c. *Go mo gaza-jec?
    2s who tell.thee-NPST.3s

d. *mo go gaza-jec?
    who 2s tell.thee-NPST.3s

Example (17) shows all the possible ways of ordering and marking the nominal constituents in a clause meaning ‘Who told you?’. The only fully acceptable utterance is (17a) which has OSV word order but follows the sequence theme-rheme. (17b) with SOV word order is intelligible but is clearly not the familiar way of saying this. From this we must conclude that the preference for the order theme before rheme may override the preference for SOV word order.

In the preceding examples we have seen that there are two principles that govern the order of the nominal constituents in a clause. The first concerns the grammatical relations and stipulates that SOV is the preferred word order. The second concerns the information structure and stipulates that theme before rheme is the preferred order. We have already seen that these principles may come into conflict with one another. Then one of the
principles — in our case the one relating to information structure — prevails over the other. However, some of the examples in (16) and (17) actually went against both of these principles; none of those examples were acceptable. That sentences which go against both principles of preferred constituent order are not acceptable can be further seen in (18) and (19).

(18) a. \( GBa\-ge\-zi\) wâtuŋ \( B\)aware-jec.
   younger.brother-2s.POSS-ERG post bring-NPST.3s
   ‘Your brother has brought the posts.’

   b. Wâtuŋ \( gba\-ge\-zi\) baware-jec.
   post younger.brother-2s.POSS-ERG bring-NPST.3s
   ‘Your brother has brought the posts.’

(19) a. \( GB\-ge\-zi\) \( WE\)mo baware-jec?
   younger.brother-2s.POSS-ERG what bring-NPST.3s
   ‘What has your brother brought?’

   b. \( ?WE\)mo \( gba\-ge\-zi\) baware-jec?
   what younger.brother-2s.POSS-ERG bring-NPST.3s
   ‘What has your brother brought?’

In (18) the direct object is a noun. Examples (18a) and (18b) show that both SOV and OSV word order are possible. In (18b) the rhematicity values of the constituents are adjusted so that the fronted direct object becomes thematic. This is not possible in (19) where the direct object is an interrogative pronoun. Since the interrogative pronoun is inherently rhematic, (19b) appears to have rheme-theme order. As this comes on top of OSV word order the sentence is judged to be of doubtful acceptability.

5 Conclusion

In this paper the conditions of use of the optional ergative of Kâte in transitive clauses were studied. We found two environments in which the use of the ergative is virtually obligatory: in clauses with the marked word order OSV, and when the subject is bound by a focus particle. These two environments have in common that the subject is highly rhematic in them. In the Kâte clause the theme usually precedes the rheme and highly rhematic elements such as interrogative pronouns are drawn to preverbal position, away from the beginning of the clause. The high rhematicity of the subject in OSV clauses falls into line with this. In quantitative analysis of discourse we found further evidence that high rhematicity triggers ergative marking. Personal pronouns, which are inherently thematic, are less likely to be marked by the ergative than nouns. Furthermore, in transitive clauses with an overt direct object NP the subject is less likely to be ergatively marked than in clauses with an ellipsed direct object. Given that high rhematicity generally correlates with the use of the ergative, it is somewhat surprising that the emphatic pronouns do not accept ergative marking. In Kâte, ergative marking of the transitive subject is a minority pattern; in discourse more than half of the transitive subjects are unmarked. There is a distinct difference in the distribution of the ergative in monologue and in dialogue, which is consistent with the claim that pragmatic factors govern the use of this marker.

Optional ergatives appear to be widespread among Trans New Guinea languages, though only a small number of detailed descriptions is available. The rule that the subject
in a double NP clause with OSV word order must be marked by the ergative case is echoed in a number of languages. Donohue (2005:183f.) shows that ergative marking is obligatory in morphologically transitive OSV clauses in Western Dani. In Korafe, the use of the optional ergative is required in OSV clauses if the object outranks the subject on the animacy hierarchy (Farr 1999:98f.) and in Fore, if both core arguments have a human referent (Scott 1986:168f.). The sensitivity of the ergative to the linear order of the arguments presumably is an indication that the information structure plays a role in its use. Indeed Farr (1999:102f.) analyses the Korafe marker as a mixture between a pragmatic marker, indicating focus, and a semantic role marker, indicating agent or force. The Kâte marker could be described in the same terms. In Fore, on the other hand, the animacy status of the subject is the major factor determining the use of the optional ergative. The ergative obligatorily occurs on inanimate transitive subjects, but is barred from occurring on personal pronouns and proper nouns (Scott 1986:170). In Korafe and Fore, therefore, both the information structure and the animacy hierarchy play a role in the use of the optional ergative, though with a different preponderance of these factors in the two languages. The conditions of use of the optional ergative in Folopa are also twofold. If the subject ranks lower on the animacy hierarchy or equals the object, then the ergative case is obligatory; otherwise it is optional (Anderson and Wade 1988:13f.). Where it is optional, the use of the ergative signals that the agent instigates and controls the action, whereas its absence signals a lack of control. Thus in Folopa, the animacy hierarchy as well as the agentivity of the subject condition the use of the ergative.

As this brief survey shows, optional ergatives are conditioned by different factors, or a different mix of factors, in different Trans New Guinea languages. It may well be that the phenomena just reported are not unitary and that they deserve to be given different names. However, in this paper I was not concerned with terminology. I stuck to the term ‘ergative’ in the description of the Kâte facts for the following reasons. First, in the overwhelming majority of its occurrences in discourse the case-marker -zi is found on the subject of a transitive verb; intransitive subjects are only occasionally so marked. Secondly, when informants are confronted with examples of a marked intransitive subject taken from discourse, they sometimes consider this problematic and suggest improving the clause by omitting the ergative marker. Marked transitive subjects never occasion such a response. Thirdly, the case enclitic -zi is not only the marker of the ergative but also of the instrumental case. Homonymy of these two cases is found very frequently in indubitable ergative languages. For these reasons it seems justified to call Kâte -zi a rhematic ergative case marker.

References


~ Complex predicates and complex sentences ~
1 Introduction

One of my most memorable classes in graduate school was a seminar that pitted a theoretical framework of passing fancy — generative semantics — against data from a descriptively intractable language of enduring fascination — Kalam. Andy Pawley served as the Kalam informant. He and Ralph Bulmer had just finished compiling a first dictionary of the language, but he had not yet published the first of many articles on aspects of Kalam syntax and discourse, including several to be referenced herein. Little did I know at the time how much I would have to confront many of the same issues that make Kalam ‘a language which defies description by ordinary means’ (Pawley 1993) when I wrote up the results of my own dissertation research on Numbami, an Austronesian language of Morobe Province, Papua New Guinea, which Andy himself — as one of the principal investigators (with George Grace) on our field research grant — had identified as conservative enough to shed light on the history of the more innovative Austronesian languages in the region.2

Comparison of an Austronesian language like Numbami with a Papuan language like Kalam reveals many more differences than similarities, of course, but the two unrelated languages are alike in relying heavily on verb serialisation to render complex events in a sequence of simplex events. In the terminology of Talmy (2000:25), verbs in both Kalam and Numbami can be said to exhibit low conflation — in other words, minimal incorporation or lexicalisation of multifaceted verbal events into individual verbs. In a series of articles over the years, Pawley has explored how the verbal system of Kalam challenges our customary approaches to translation (1991), grammatical description (1993), and lexicography (2006), while Lane (1991) has described Kalam serialisation in considerable detail. My own output on the verbal system of Numbami has so far been limited to its history (1982) and the subject relationships that obtain within serial verb constructions (1993, 1999). However, Pawley and

---

1 I am indebted to Yuko Otsuka and to an anonymous referee for very helpful criticism of an earlier draft of this paper.

2 Andy proved right, though not always in the domains he had anticipated. Several key elements of reconstructed Oceanic morphology, in particular, seem to have gone missing in Numbami. The University of Hawaii Oceanic Comparative Linguistics Project directed by George Grace and Andrew Pawley was funded in part by National Science Foundation Grant No. BNS 75-19451. George later served as my dissertation advisor.
Lane (1998) provide a starting point for extending the analysis to the nature of event construction in certain kinds of serial verb constructions (SVCs), in other words, how simplex events combine to form complex events.

Pawley and Lane (1998) discuss discourse formulas that build what they call ‘multi-scene SVCs’—events that start in one place and continue in another. For instance, they provide the following formula for the ‘core elements of deliberate action event reports’: (a) MOVE/STAY, (b) PIVOTAL ACT, (c) MOVE (WITH AFFECTED OBJECT), (d) OUTCOME (1998:212). They note the unique combinatorial privileges of ‘orientation verbs’ (‘go’, ‘come’, ‘ascend’, ‘descend’, ‘walk’, etc.) in such SVCs, where they tend to occur before the PIVOTAL ACT and/or before the OUTCOME.

Now, Kalam is a clause-chaining, switch-reference language, and Pawley and Lane focus particular attention on the fuzzy boundaries between multi-scene SVCs and multi-clause chains. My goal here is to investigate in more detail (to ‘home in on’, to use an orientation metaphor) the internal semantic components of the motion events themselves in Numbami, not Kalam. Neither the semantic label MOVE nor the descriptive label ‘orientation verb’ does justice to the range of verbal elements that can combine to build complex motion-event SVCs out of semantically simpler lexical components in serialising languages like Numbami. This deconstruction of motion-event constructions in Numbami will illustrate one of the most often quoted characterisations of serialised verbs as describing ‘subparts or aspects of a single overall event’ (Lord 1973:269).

2 Components of motion events

There is a vast literature on motion verbs and motion events. Frawley (1992:172) lists eight components of motion events (slightly reworded here), not all of which are separately encoded in each language:

- the thing displaced: theme or figure
- the starting point: source
- the destination: goal
- the trajectory: path, direction
- the location: site
- the means: instrument or conveyance
- the manner of movement
- the cause: agent

Path involves three of these components: starting point, trajectory, and destination. Manner of movement is of related interest because the lexical encoding of manner and path yields an interesting typology of motion events first proposed by Talmy (1985), and later revised and elaborated (Talmy 2000). The seminal insight can be summarised thus. Languages such as Spanish and Japanese tend to encode path in the main verb when describing motion events, while relegating manner to a satellite role, expressed, for example, by gerundive constructions. An English analogue would be ‘They entered the house running’. In contrast, languages such as German or Finnish tend to encode manner in the main verb, while relegating path to a satellite role, expressed, for instance, by adpositional phrases. An English analogue would be ‘They ran into the house’.
Huang and Tanangkingsing (2005) have proposed adding two additional categories to this typology, after carefully examining reference to motion events in six Austronesian languages that represent four different primary branches of the language family: Cebuano, Malay, Saisiyat, Squliq Atayal, Tagalog, and Tsou. While acknowledging that most of their sample exhibits ‘unmistakable strong path salience’ when classified according to Talmy’s original two-way typology, they note that Tsou has developed verbal compounds consisting of a manner prefix and a path root ‘that give equal salience to path and manner’ (2005:311). Perhaps the same can be said for the many Oceanic languages in New Guinea (like Manam) and New Caledonia (like Tinrin) that exhibit complex predicates consisting of a manner (‘classifier’) prefix and a path root. ‘Macro-events’ is the label Huang and Tanangkingsing use for these complex predicates. Serialisation is the final category in their four-way typology, summarised here.

- **Satellite-framed language**: Path satellite + Manner verb
- **Verb-framed language**: Path verb + Manner adjunct
- **Macro-event language**: [Manner prefix + Path root]VERB
- **Serial verb language**: Manner verb + Path verb

Although useful for distinguishing Tsou from the other languages in their sample, the latter two categories in this typology conflate two variables: manner or path salience and clause structure. Huang and Tanangkingsing (2005:310) note that the order of manner and path constituents is not crucial, but this is harder to construe in the case of the ‘macro-event’ construction. A more order-neutral formulation would allow for [Path prefix + Manner root] or [Manner root + Path suffix]. While clause structure is the key determinant of whether a language is satellite-framed or verb-framed, languages with complex predicates or verb serialisation can be either satellite-framed, verb-framed, or equipollently framed (Slobin 2004).

When both manner and path are rendered by full verbs in serial verb constructions, one is tempted to classify SVCs as equipollent, privileging neither manner nor path, just as in Tsou verbal compounds. However, Huang and Tanangkingsing (2005) note several asymmetries. In path-oriented serialising languages one might expect path verbs to precede manner verbs, while manner verbs precede path verbs in manner-oriented languages, but in their small corpus of SVCs in western Austronesian languages, Huang and Tanangkingsing found that path verbs never precede manner verbs in motion-event SVCs, except in constructions of the type ‘go and V’ or ‘come and V’ (2005:318). In the few cases where both verb types do co-occur, the order is manner + path. Nevertheless, in path-oriented languages, manner verbs remain optional in motion-event SVCs, while path verbs are obligatory. Although my Numbami data accord with Huang and Tanangkingsing’s observations with regard to the order of constituents, I do not have access to a controlled corpus such as a frog story narrative (see Slobin 2004). This study will therefore limit itself to investigating the micro-components that are used to construct macro-events.

The verbs that participate in Numbami manner and path serialisation can be classified into four categories, each class represented by a Numbami prototype verb, or what Talmy (2000:37) calls a mid-level morpheme, as opposed to a deep morpheme.

---

3 Talmy (2000:49) lists Polynesian languages among those that resemble Spanish and Japanese in encoding path in main verbs and manner in satellite roles.
A deep morpheme represents a concept that is believed to be both fundamental and universal in the semantic organization of language. A mid-level morpheme represents a particular conceptual complex that consists of a deep-morphemic concept together with certain additional semantic material, and that is recurrent in the semantic organization of a particular language, though it is often also to be found in many other languages. [emphasis in original]

In Numbami, each category of verb adds a complementary semantic component within SVCs representing motion events: MOVE verbs add a manner of motion, such as walking, running, flying, sailing; GO verbs add deictic directions such as toward the speaker, toward the addressee, or away from both; AIM verbs add a trajectory, such as up or down, back or forth, on or off; while REACH verbs (or prepositions) add goals or destinations marking endpoints of motion. Other verbal prototypes can combine with the path elements to form even more complex motion events. These include HOLD verbs (e.g., ‘hold, carry, take, transport’), which add more about either theme or instrument; PUT verbs (e.g., ‘put, give, appoint, send’), which add more about the cause of motion; and STAY verbs (e.g., ‘dwell, lie, sit, stay’), which add more about the location or duration of the motion events. In addition, ACT will serve as a catchall label for other sorts of action verbs that may participate in SVCs containing motion events. Examples (1) through (4) illustrate various combinations of these classes of verbs.

(1) MOVE AIM GO
Balus i-lowa i-leleu i-ma.
airplane 3s-fly 3s-return 3s-come
‘The airplane flew back here.’

(2) MOVE AIM GO REACH
Ma-paandalowa ma-woti ma-ma ma-solonga teteu.
1pe-walk 1pe-descend 1pe-come 1pe-enter village
‘We walked down here into the village.’

(3) AIM GO ACT
Mana-pi mana-uwa mana-ani bani?
1pe.IRR-ascend 1pe.IRR-go.to.2 1pe.IRR-eat food
‘Are we to come up (to where you are) and eat?’

(4) ACT GO REACH
Na-so pepa ni-wesa ni-ndenga nanggi lunggewe.
IRR.1s-stab paper IRR.3s-go IRR.3s-reach my sister
‘I’ll write a letter to my sister (far away).’

3 Manner and path verb classes in Numbami

3.1 Move verbs: manner of motion

Like Spanish or Japanese, Numbami does not conflate manner-of-motion with path-of-motion in a single verb. To get manner and path in the same clause, Numbami needs both a manner verb and a path verb in a serial construction; it has no means of deriving gerundive or conjunctive (or even nominalised) forms from verbs, as Japanese does. Compare the Japanese (from Shibatani 2003:264) and Numbami sets of examples in (5) and (6), respectively. The highly unnatural sentences (5a) and (6a) specify both manner of motion and destination, but not path of motion. The addition of path verbs in (5b) and (6b) makes the sentences sound much more natural.

(5) a. MOVE
   *Ken-wa gakkoo-ni arui-ta.
   ‘Ken walked to school.’

   b. MOVE GO
   Ken-wa gakkoo-ni arui-te it-ta.
   lit. ‘Ken went to school walking.’

(6) a. MOVE
   *Ekapa-kolapa ti-dodomu su lumana.
   ‘The children ran to school.’

   b. MOVE GO
   Ekapa-kolapa ti-dodomu ti-wesa su lumana.
   ‘The children ran off to school.’

3.2 Go verbs: deictic direction

Numbami deictic verbs distinguish three directions: -ma ‘come toward speaker’, -uwa ‘go toward addressee’ (glossed here as ‘go.to.2’), and -wasa ‘go away from either speaker or addressee’. These verbs are ubiquitous in Numbami discourse, although -uwa ‘go toward addressee’ is by far the rarest of the lot. Not only do they cover the functional range of ‘come’ and ‘go’ in most other languages; they also add directionality to manner-of-motion (MOVE) verbs, and deictic directionality to other directional (AIM) verbs. Finally, they also perform functions similar to directional adverbs such as here and there in English (or hither, thither, hence, thence, and yonder in more archaic English).4

4 Jabêm also exhibits a three-way distinction in its deictic verbs (see Bisang 1986:139; Bradshaw and Czobor 2005:19–20). The initial vowel of the verb stem -wasa assimilates to the vowel of its subject prefix, thus: wa-wasa ‘1s-go’, u-wosa ‘2s-go’, i-wesa ‘3s-go’.

5 Lichtenberk (2003) describes an unusual division of directional duties in Toqabaqita, in which there is no simple verb ‘come’, while the verb ‘go’ lacks any deictic orientation. Instead, two deverbal directional particles, andative kau and venitive mai, do most of the deictic work, and much else besides.
Although deictic directional verbs in many languages are intransitive, Numbami go verbs can take overt direct objects, so long as (a) those objects indicate target locations, and (b) those target locations are compatible with the deictic target direction of each verb: toward speaker, toward addressee, or away from either. Numbami makes no morphological distinction between transitive and intransitive verbs.

Another important point to note about go verbs in Numbami is that they denote movement toward a target location, but make no claim about arrival at that target. Thus, ti-wesa Lae (lit. ‘they went Lae’) is more precisely translated ‘They left for Lae’ or ‘They went Lae-ward’ instead of ‘They went to Lae’. The presence of a reach verb or preposition is required to specify arrival at the endpoint of a path. Example (10) illustrates this difference, and also suggests why most go verbs with overt objects are marked for irrealis, marking travel not yet begun. When reporting past travels, we are more likely to know whether or not the travellers reached their destinations.

3.3 Aim verbs: trajectory of motion


Although deictic directional verbs in many languages are intransitive, Numbami go verbs can take overt direct objects, so long as (a) those objects indicate target locations, and (b) those target locations are compatible with the deictic target direction of each verb: toward speaker, toward addressee, or away from either. Numbami makes no morphological distinction between transitive and intransitive verbs.

Another important point to note about go verbs in Numbami is that they denote movement toward a target location, but make no claim about arrival at that target. Thus, ti-wesa Lae (lit. ‘they went Lae’) is more precisely translated ‘They left for Lae’ or ‘They went Lae-ward’ instead of ‘They went to Lae’. The presence of a reach verb or preposition is required to specify arrival at the endpoint of a path. Example (10) illustrates this difference, and also suggests why most go verbs with overt objects are marked for irrealis, marking travel not yet begun. When reporting past travels, we are more likely to know whether or not the travellers reached their destinations.

Although deictic directional verbs in many languages are intransitive, Numbami go verbs can take overt direct objects, so long as (a) those objects indicate target locations, and (b) those target locations are compatible with the deictic target direction of each verb: toward speaker, toward addressee, or away from either. Numbami makes no morphological distinction between transitive and intransitive verbs.

Another important point to note about go verbs in Numbami is that they denote movement toward a target location, but make no claim about arrival at that target. Thus, ti-wesa Lae (lit. ‘they went Lae’) is more precisely translated ‘They left for Lae’ or ‘They went Lae-ward’ instead of ‘They went to Lae’. The presence of a reach verb or preposition is required to specify arrival at the endpoint of a path. Example (10) illustrates this difference, and also suggests why most go verbs with overt objects are marked for irrealis, marking travel not yet begun. When reporting past travels, we are more likely to know whether or not the travellers reached their destinations.
(12) **AIM** **GO** **ACT**

Ma-pi **ma-wasa** ma-yanggo **bingsu** Lene.
1pe-**ascend** 1pe-**go** 1pe-**see** missionary Lehner
‘We went up and saw missionary Lehner.’

(13) **MOVE** **AIM**

Ma-kota tina ma-sakiya teulu.
1pe-**wade** river 1pe-**embark** side
‘We waded the river and climbed up the other side.’

(14) **AIM** **REACH**

Wangga i-supula bubusu i-solonga molou.
canoe 3s-**round point** 3s-**enter cove**
‘The canoe rounded the point into the cove.’

Like the GO verbs, these AIM verbs can take transitive objects, so long as those objects denote locations appropriate to the trajectory. These would denote target locations in the case of -kawewe ‘steer, turn toward’ or -sakiya ‘embark, climb up onto’; source locations in the case of -kosa ‘disembark, climb down from’ or -weke ‘leave, abandon’; and site (or ground, as opposed to figure) locations in the case of -pi ‘ascend, climb up’ or -supula ‘round (a point)’.

### 3.4 Reach verbs: goal or destination

The roots of the two specialised REACH verbs in Numbami double as prepositions. Other verbs of motion can serve as REACH verbs when they occur at the ends of path constructions, as in (14), but the main focus here is the specialised members of the set. When the REACH component of a motion event is represented by a preposition rather than its corresponding verb, the resulting construction may still be considered a path construction, even though it may not be considered an SVC unless it also contains at least two inflected verbs.

The path verb -su(wa) ‘reach; arrive at, onto, into’ is matched by the more general locative/goal preposition su(wa) ‘at, onto, (up)on, to’. The latter is far more common. In fact I have no examples of the verb except in path constructions. Both the verb and the preposition take locative objects, but the object of the verb designates only a locative goal, as in (15), while the object of the preposition can designate either a locative goal, as in (17), and the first clause of (16), or the location or time of an event, as in (18) and the second clause of (16). Either the verb or the preposition can occur at the end of path constructions, but in either case its object must designate a goal, not a static location.

---

6 The same root shows up in the lexical compound -ndosuwa ‘sit down’, from -ndo ‘stay, sit’ + suwa ‘onto, upon’. The compound is used intransitively, although of course it implies a locative goal, a place to sit down upon. The suwa in the compound is not truncated to su, as the preposition usually is. The compound -ndosuwa is thus distinguishable from the sequence of -ndo + su(wa), as in -ndo su zamoka ‘sit on veranda’.
 MOVE STAY REACH
Wa-dodomu wa-mi wa-su nanggi kapala.
1s-run 1s-stay 1s-reach 1s.POSS house
‘I ran on down to my house.’

GO REACH + ACT ON (LOCATION)
Wa-wasa wa-su dume wa wa-walasa su wati.
1s-go 1s-reach back and 1s-stand upon rock
‘I went around the back and stood up on a rock.’

GO TO (GOAL) [...] REACH
Ma-wasa su tina ti-kamba ti-nggo Busu. Ma-suwa nato, tako, ...
1pe-go to river 3p-call 3p-say Busu 1pe-reach there okay
‘We went as far as the river they call the Busu. We arrived there, okay, …’

ACT AT (TIME)
Ti-solonga lumana su ait kilok.
3p-enter school at 8 clock
‘They started school at 8 o’clock.’

REACH
Tiyamama ti-su weni wai.
everyone 3p-into bush CMP
‘Everyone had gone into the forest.’

The rather rare path verb -ndenga ‘reach; arrive at’ is matched by the far more common generalised dative preposition de(nga) ‘to, at’. The verb appears limited to path constructions, as in (4), repeated here as (20). Both the verb and the preposition take dative goals as their object complements.

ACT GO REACH
Na-so pepa ni-wesa ni-ndenga nanggi lunggewe.
IRR.1s-stab paper IRR.3s-go IRR.3s-reach my sister
‘I’ll write a letter to my sister (far away).’

GO TO (GOAL)
Gaya nu-ma denga woya.
tomorrow IRR.2s-come reach 1s.O
‘Come find me tomorrow.’

SAY TO (DATIVE) SAY
Lawa ata ina-nggo binga de ata, ina-nggo ka ... 
people later 3p.IRR-say talk to self 3p.IRR-say like
‘People will then talk to each other, they’ll say (like) …’

---

7 My impression is that the preposition denga ‘to’ is most often truncated to de, while the much rarer verb -denga ‘reach’ is never shortened. In this respect, the verb–preposition pairs -denga ‘reach (a dative goal)’ and de(nga) ‘to’, and -su(wa) ‘reach (a locative goal)’ and su(wa) ‘at, onto, (up)on, to’ are not parallel.
Manner and path verb serialisation in Numbami

(23) PUT TO (DATIVE)
    \[ E \ i-ki \ danke \ deng \ i-lu \ \text{wai}. \]
    3s 3s-put thanks to 1pe-two CMP
    ‘He conveyed his thanks to the two of us.’

The same root -ndenga appears in the multifunctional verb -ndengama ‘reach, match, suffice; be possible’, often intertranslatable with Tok Pisin inap. This verb can only be inflected for third person, and its most common uses are in abilitative constructions such as (24), where it takes a sentential complement, or as the equivalent of ‘okay, then’ as in (25). It also introduces an adverbial phrase with an object complement containing, for instance, a specific number or time, as in (26) and (27). (In fact, the final -ma on this form may reflect the Numbami adverbial marker.) Since the verbal categories of motion event construction are not at issue here, the superordinate prototype verb labels employed above will be omitted from (24–27).

(24) Woya i-ndengama kote inggo na-tamu.
    1s 3s-suffice not CMP IRR.1s-join
    ‘I can’t join you.’

(25) I-ndengama go, ta-wasa!
    3s-suffice then 1pi-go
    ‘Okay, then, let’s go.’

(26) Nu-ambi buwa ni-ndengama ai-to toli.
    IRR.2s-hold betel.nut IRR.3s-suffice 1pi-few three
    ‘Get betel nuts enough for the three of us.’

(27) Gaya, yawela gedo i-ndengama twelv kilok,
    tomorrow sun shadow 3s-reach 12 o’clock
    nu-leleu nu-ma.
    IRR.2s-return IRR.2s-come
    ‘Come back tomorrow at 12 noon.’

3.5 Wider implications: motion event construction

What larger implications can we draw from these deconstructed representations of complex events? First, it seems clear that Pawley and Lane’s (1998:212) ‘core element’ MOVE, which occupies two separate slots in the discourse templates for ‘deliberate action event reports’ in Kalam, is itself decomposable into smaller elements: manner of motion (MOVE), trajectory of motion (AIM), deictic direction (GO), and goal or destination (REACH). Second, the order of these elements in Numbami accords with Huang and Tanangkingsing’s (2005:318) observation about their small sample of motion-event SVCs in Western Austronesian languages, namely, that path verbs never precede manner verbs, except in constructions of the type ‘go and V’ or ‘come and V’. Moreover, Numbami

---

8 Jabêm has a cognate adverbial verb form -(n)dêŋ ‘move toward a goal; reach; to, at; when’, which likewise can only be inflected for third person (gêdêŋ/êndêŋ) and has a wide range of functions, among them, introducing datives, specific times, and time clauses ( Bradshaw and Czobor 2005:41–42, 48–49, 112).
manner + path SVCs appear to resemble Tsou verbal compounds in being equipollent, privileging neither manner nor path where both occur in motion-event constructions. Finally, motion-event constructions in Numbami seem to present prime examples of one of the earliest (in English) and most often quoted characterisations of serialised verbs as describing ‘subparts or aspects of a single overall event’ (Lord 1973:269).

## 4 Postscript: combinability and polysemery

Pawley’s encounter with Kalam has spurred him to articulate several special challenges presented by languages whose verb roots exhibit low conflation, to use Talmy’s (2000:25) term, or — to use a different metaphor — languages whose speakers might be said to construct verbal narratives out of elements closer to semantic atoms than to semantic molecules.

Pawley (2006) also observes that verbs in any language that participate in a multitude of combinations tend to exhibit an extraordinarily high degree of polysemy. This is certainly true of Numbami. Take, for instance, the class of verbs labelled PUT in the discussion above of prototypes that often combine with path verbs. In fact, if \( \text{Displace}(x) \) is ‘the elemental structure of a motion event’ (Frawley 1992:171), then PUT is its transitive counterpart, \( \text{Displace}(x,y) \). The semantics of the PUT class ranges from ‘put’ to ‘send’ to ‘give’ to ‘appoint’, each of which is rendered in Numbami by the same verb -\( ki \) in combination with different elements.

(28)  **PUT ON (PLACE)**

\[
\text{Ina-}ki \quad \text{kundu na ano suwa nato.}
\]

3p.IRR-put sago 3s fruit onto there

‘They’ll put the starch from the sago onto it.’

(29)  **PUT TO (PERSON)**

\[
\text{Ti-}ki \quad \text{bani de e kote.}
\]

3p-put food to 3s not

‘They didn’t give him any food.’

(30)  **PUT GO**

\[
\text{Wa-ki} \quad \text{biya i-uwa kote.}
\]

1s-put beer 3s-go.to.2 not

‘I didn’t send you any beer.’

(31)  **PUT AIM GO**

\[
\text{Bingsu i-ki binga i-woti i-wesa.}
\]

missionary 3s-put word 3s-descend 3s-go

‘The missionary sent word down (and out).’

(32)  **PUT (PERSON)**

\[
\text{Ina-}ki \quad \text{lawa lomosanga na wa lawa ina-lapa na.}
\]

3p.IRR-put people leaching of and people 3p.IRR-beat of

‘They’ll appoint people for leaching and people to pound.’

When combined with REACH and a locative goal, as in (28), -\( ki \) translates as ‘put’. When combined with REACH and an animate goal as in (29), -\( ki \) translates as ‘give’. When combined with AIM or GO, but not REACH, as in (30)–(31), -\( ki \) translates as ‘send’. When
the direct object of put denotes one or more persons, and there are no subsequent path elements, as in (32), -ki translates as ‘appoint’ or ‘nominate’. Are there four separate lexemes -ki (‘put’, ‘give’, ‘send’, ‘appoint’) in Numbami, each taking a separate very broadly defined combination of path elements (or none at all) as complements? Or is there one semantically consistent -ki (‘PUT’) that can play a role in at least four broad types of events (placing, giving, sending, and appointing)? In English, do we need to recognise at least two separate verbs for send and give and put in order to account for the respective differences between send me a letter and send away for a prize, between give me money and give away a prize, between put me to sleep and put a letter in the mail? Or is it sufficient to recognise that the semantics of prototype put verbs — which can signal initiation of a transfer of position, status, ownership, or even sensation — are compatible with a very wide range of different results, perhaps especially where motion along a path is involved?

Similar questions apply to other types of verbs in event constructions. Consider a prototype verb that could perhaps be named feel. The form -mandi ‘fear’ can combine with either nothing, translating ‘be afraid’; or a direct object, translating ‘be afraid of, fear, flee (something)’; or a path, in which case it translates as ‘flee (to)’. Are there three verbs -mandi or is there a consistent emotion that can show up in three different event types? Inasmuch as Numbami (a) makes no morphological distinction between transitive and intransitive verbs, and (b) allows zero pronouns in object position, it is often hard to tell without rich discourse context whether a particular instance of timandi should be translated as ‘They’re afraid’, ‘They’re afraid of it’, or ‘They flee’.

\[(33) \text{ FEEL} \]
\[
\begin{array}{ll}
\text{Aiya u-mandi.} & 2s \text{ 2s-fear} \\
\end{array}
\]

‘You’re afraid (of it).’

\[(34) \text{ FEEL (THING)} \]
\[
\begin{array}{ll}
\text{Kolapa kae i-mandi bumewe.} & \text{boy little 3s-fear whites} \\
\end{array}
\]

‘The little boy is afraid of white people.’

\[(35) \text{ FEEL REACH GO} \]
\[
\begin{array}{ll}
\text{Ti-mandi su weni ti-wesa.} & \text{3p-fear to bush 3p-go} \\
\end{array}
\]

‘They fled into the forest.’

It may be that highly polysemous, highly recombinant, but relatively simplex predicates in serialising languages like Numbami and Kalam combine in relatively egalitarian fashion, imposing relatively few selectional restrictions that extend beyond each atomic verb phrase. Perhaps our view of the verbal clause as composed of a dominant verb that selects for its clausal complements but not its clausal adjuncts reflects our own greater familiarity with languages that conflate a wider range of simplex elements into each lexical verb. Clauses in languages like Numbami and Kalam seem simultaneously less hierarchical and more structurally flexible in the hands of accomplished storytellers in such languages than do clauses in languages like English, with its huge number of lexical verbs conflating more complex notions and imposing greater order on their surrounding clauses.
References


Event integrity and continuity in the Tuam language

ROBERT D. BUGENHAGEN

1 Introduction

One of the principle defining characteristics of well constructed texts in any language is that they exhibit a high amount of coherence. This is defined by Givón (1995:343) as follows:

Coherence is the continuity or recurrence of some element(s) across a span (or spans) of text.

Of the many types of elements that can recur across spans of text and provide continuity, Givón (1995:343) lists the following six as being among the most concrete and easiest to track: (1) referents, (2) temporality, (3) aspectuality, (4) modality/mood, (5) location, (6) action/script. The present paper focuses on the last of these, describing the encodings of events in narratives in the Tuam dialect of the Saveeng (also referred to in the literature as Mutu or Mutu-Tuam) language, focusing particularly on connective devices and how they reflect varying degrees of continuity of action between the successive events recounted in a narrative.¹

The Saveeng language is an Oceanic Austronesian language spoken by some 3500 people living on Umboi Island and on the nearby small Siassi Islands in the Morobe Province of Papua New Guinea. There are three dialects in the language: Tuam, Oov (sometimes referred to as ‘Mutu’ in the literature), and Malai. Ross (1988:122) classifies Tuam as a member of the Tuam network, in the Ngero [sub-RDB] Family and Ngero-Vitiaz Family within his North New Guinea Cluster.

Typologically, Tuam is a head first language with subject–verb–object–prepositional phrase word order, prepositions, and complementisers which precede the subordinate clauses they govern. Most modifiers in the noun phrase occur after the head noun, and there is a distinction between alienable and inalienable genitives. Inalienable genitives and certain ‘topicalised’ alienable genitives precede the nouns with which they are associated.

¹ The author and his wife have been resident in the Tuam-speaking village of Yaagha and conducting research and facilitating translation and vernacular literacy there since April 2000, working under the auspices of the Summer Institute of Linguistics. In the orthography, nearly all symbols have their normal phonetic values. Gh represents a voiced velar fricative, v a bilabial fricative, and doubled vowels (aa, ee, ii, oo, uu) represent long vowels. I would like to acknowledge the help given to me by Malcolm Ross. His comments and suggestions certainly helped to improve the paper.
Non-topicalised alienable genitives occur following the head noun. Verbs are inflected with subject-indexing prefixes, are reduplicated to overtly indicate imperfect aspect, and may undergo several different types of transitivity derivation including: (1) causative formation, (2) detransitivisation, and (3) applicative formation. The following four examples illustrate most of these features.

(1)  
\textit{Yau na-ghol ŋgai tintiina ru tane.}  
\textit{1s.NOM 1s-buy pig big:RDP two this}  
‘I bought these two pigs.’

(2)  
\textit{Ta-gham moon raau pa atov to ruum.}  
\textit{1pi-do/get sago leaf:3s.GEN OBL thatch LOC house}  
‘We (IN) use sago leaves for the thatching of the house.’

(3)  
\textit{Yau na-saav pa sav-e-ŋŋ tau noor i}  
\textit{1s.NOM 1s-speak OBL speak-SV-NR that yesterday 3s.NOM}  
\textit{i-sav-i-a na.}  
\textit{3s-speak-APP-TR GIVEN}  
‘I spoke about what he said yesterday.’

(4)  
\textit{Maet i-pa-pa-burig-in ŋgar sa-saghat-i naol.}  
\textit{stone/money 3s-RDP-CAUS-get.up-APP thinking RDP-be.bad-N.ATTR much}  
‘Money causes much bad thinking/behaviour.’

The Saveeng language exhibits a rich array of intrasentential connective devices for encoding events in narratives, whose use seems to be mainly conditioned by varying degrees of \textit{continuity} between the events being related. In the author’s experience, learning to correctly use these devices presents a considerable challenge for the outsider who wants to learn to speak the language. This paper presents a survey of these various devices and attempts to explicate the notion of \textit{continuity of events} and apply this to distinguishing the functions of the various devices. The devices are listed below.

1. serial parataxis of clauses under a single intonational contour with no intervening conjunctions or pauses (i.e. serial verb constructions)
2. linkage of clauses with the conjunction \textit{gha} ‘and’
3. linkage of clauses with the telic conjunction \textit{le} ‘until’, ‘and as a result’
4. co-ordinate linkage with the conjunction \textit{ve} ‘and’ or co-ordinate parataxis (with rising intonation at ends of clauses and often pauses between clauses)
5. linkage with the conjunction \textit{ra(gha)} ‘(and) then’

2 The devices examined in detail

2.1 Serial parataxis/verb serialisation

In serial parataxis the components of the serialisation construction all occur under a single intonational contour with no intervening conjunctions or pauses. The conjuncts in such constructions typically exhibit sharing of operators for modality and negation.

(5)  
\textit{I i-rau [i-gharaat sav-e-ŋŋ tane i-dudun].}  
\textit{3s.NOM 3s-able 3s-fix speak-TR-NR this 3s-be.right}  
‘He is able to fix this matter properly.’
(6) [Ti-gham ŋgar eta i-sosor] mako.
   3p-do/get thinking/behaviour any 3s-be.wrong not
   ‘They did not do anything wrong.’

Core arguments are also shared among the linked predicates and are normally all represented in the first component of the serialised construction. The notion of what constitutes a core argument is somewhat problematic, however. Subjects and objects are obviously core arguments, but many prepositional phrases and their pronominal equivalents seem to function as ‘oblique’ arguments. Although they are optional, their semantic role varies significantly according to the verb or non-verbal predicate with which they occur. For example, although prepositional phrases headed by the all-purpose oblique preposition pa express adjunct roles like time, reason, and outer location, they also exhibit such semantic roles as source, addressee, benefactive, and malefactive, depending on the predicate with which they cooccur. Because of this tight semantic connection with the predicate, they are somewhat like arguments. Such ‘oblique’ arguments are also included when considering this matter of argument sharing in verb serialisation constructions. As one example, a very common construction is: [-saav ‘speak’ pa NP.addressee Verb…], where the subject of the following verb is coreferential with the addressee. The whole construction is pronounced under a single intonational contour with no pause before the following verb and expresses the meaning ‘tell X to do Y’.

(7) A-saav pani i-lam.
   2p-speak 3s.OBL 3s-come
   ‘Tell him to come.’

However, there are also some instances of what Crowley (1987:48) terms ‘ambient’ serialisation, in which a serialised clause semantically modifies the whole predication of the preceding clause rather than applying to a particular referent.

Ambient serialisation … is the term used to refer to a construction in which a verb is serialised to another verb but in which there is no specific referent associated with the subject of the serialised verb, and the verb simply describes a general predication.

Examples (5) and (6) above are instances of ambient serialisation.

Paratactic serialisation has a number of quasi-grammatical functions, expressing certain semantic case roles, aspect, and quantification, as well as also introducing various types of quotative complements. It also is used to express extremely common, almost formulaic series of events.

2.1.1 Quasi-grammatical uses of serial parataxis

Motion verbs are the most commonly occurring verbs in serialised constructions. They are used to express inner locative arguments and a number of different adverbial adjuncts functioning as semantic modifiers of the clausal core.

1. Inner locative arguments — locations where entities are asserted to be, and locative goals and sources.

(8) Dgavuun i-la i-leep i-zi nughei.
   dog 3s-go 3s-stay 3s-descend village
   ‘The dog went and stayed in the village.’
(Inner location. The verb -la is used with smaller more well-defined locations, while -zi ‘descend’ is used with larger areas or less precise locations. Location on top of something is frequently expressed using the motion verbs -zaa or -zala ‘ascend’. Location inside of something is expressed using the motion verb loon ‘enter’ or the sequence -looŋ-la ‘enter go’. Note that the third person singular form of the verb normally occurs in locative serialisations.)

(9) Ti-liŋ te rigmon i-la uur.
3p-pour salt(water) little 3s-go pot
‘They poured a little saltwater into the pot.’ (Goal)

(10) Ni-ghol taan tane i-la to yes Mangaav-a.
1pe-buy land this 3s-go LOC PL Mangaav-PL
‘We (EX) bought this land from the Mangaava people.’
(Source. The verb -la ‘go’ is the only verb occurring in a serialised construction expressing the semantic role of source.)

2. Outer locative adjuncts — the location in which a situation or event takes place.

(11) Pataŋ-i eez i-vot pani i-zi eez livuugha.
be.heavy-N.ATTR one 3s-happen 3s.OBL 3s-descend road middle:3s.GEN
‘A problem happened to him in the middle of his journey.’

3. Directional adverbials (expressed by serialised motion verbs).

(12) Aniiŋ to uum ti-tum poi ti-zaa.
food LOC garden 3p-grow well 3p-ascend
ve ti-ghur anya-zi katini.
and 3p-put fruit/nut/edible.part-3p.GEN much
‘The food [plants] of the garden grew up well and produced much food.’

4. Durative aspect — expressed by a following ambient serialisation with -la ‘go’.

(13) Ti-zeuzeu i-la i-la le ravrav i-zi.
3p-forage.on.reef 3s-go 3s-go TEL afternoon 3s-descend
‘They kept on foraging on the reef until afternoon came.’

5. Geographic extent and temporal duration expressed using the verb -rau ‘able, adequate, should, extend’.

(14) Varu-zi i-laan i-rau nugh tane wa.
news-3p.GEN 3s-run 3s-extend place this PF
‘The news about them has already spread throughout this area.’

(15) I-moroo i-rau ndaman ru, ragha i-maat.
3s-sick 3s-extend year two then 3s-die
‘He was sick for two years and then died.’

6. Similarity and manner expressed by an ambient serialisation using the verb -min ‘become’, ‘be like’.
Event integrity and continuity in the Tuam language

(16) *I-sav-i-a sav-e-ey tana i-min tov-a-aŋ-ni.*
3s-speak-APP-TR speak-TR-NR that 3s-be.like test-TR-NR-3s.ACC
‘He said that as a way of testing her.’

The same verb *-min* is also used with numeric² quantifiers to express ordinal notions.

(17) *I-poop naatu ite i-min ru (pani).*
3s-give.birth.to child:3s.GEN other 3s-become two (3s.OBL)
‘She gave birth to her second son.’

The third person singular form of *-min* also functions as a counterfactual modality adverbial.

(18) *I-min ta a-loo ŋ ghau, kanaŋ pa.taŋan-i i-ndeeŋ gham mako.*
3s-be.like that 2p-listen 1s.ACC would be.heavy-N.ATTR 3s-find
gam mako.
2p.ACC not
‘If you (PL) had listened to me, then you (PL) would not have got into trouble.’

7. A serialisation construction with the verb *-taghon* ‘follow’ is used directionally to express motion along a path and more abstractly to express that something accords or fits with the standard expressed by the following noun phrase. Unlike most other instances of quasi-grammatical serialisation, *taghon* is not normally inflected when it expresses motion along a path. Such instances could also be treated as being de-verbal prepositions.

(19) *Ti-lagh-laagh taghon ee z ti-la.*
3p-RDP-walk follow road 3p-go
‘They kept walking along the road.’

(20) *Gabua i-vot i-taghon ŋgar tsiau mako.*
thing 3s-happen 3s-follow thinking 3s.LOC not
‘Things didn’t happen as I planned.’

8. The complements of most verbs of speech, emotion, and cognition occur in verb serialisation constructions introduced by the verb *-ghe* ‘say, think, want, about to’.

(21) *Ra ŋerek i-saav pa mbiaŋ i-ghe:*  
then parrot 3s-speak OBL bat 3s-say
[“Yo ana-m u-tooov-a u-tuuk.”]
2s.NOM turn-2s.GEN 2s-try-TR 2s-hang
‘Then the parrot said to the bat, “It is your (SG) turn to try and hang [upside down].”’

(22) *Yau na-watatgh na-ghe: [i loolo payau].*
1s.NOM 1s-know 1s-think 3s.NOM insides:3s.GEN 1s.OBL
‘I know that he likes/loves me (lit. ‘his insides [are] for me’).’

---
² The notion of ‘first’, however, is expressed using the form *mata-* ‘eye, sharp, first’ rather than the numeral ‘one’ *ee(z).* Compare *Itiir gabua toozi imin maata.* ‘He examined their things first.’
(23) *Dgar tooni i-yaryaan i-ghe:* [i-mogheer tomani zi].

thinking 3s.LOC 3s-strong 3s-want 3s-play with 3p.ACC

‘He badly wanted to play with them.’

Note that if there is a pause, it would hardly ever occur before the verb *ghe*. Instead, it follows *ghe*, coming before the complement clause it governs.

*Ghe* also functions as a main verb expressing the notions ‘say’ and ‘want’. And the notion of ‘if’ in hypothetical conditionals is expressed by the third person singular form of *ghe* in optional combination with the third person singular form of *saav* ‘speak’: (*isaav*) *ighe*.

(24) * (Isaav) *ighe* yo *u-watagh* bod-e-ey,

if 2s.NOM 2s-know write-TR-NR


‘If you (SG) know how to write, then you (SG) can write letters to your (SG) friend.’

9. A final, extremely common quasi-grammatical use of paratactic serial clauses involves the use of the verb *-sov* ‘end, finish, all’ to express completive aspect (25) or universal quantification ‘all’ (26).

(25) *Ti-gham narogho tana i-sov,* ve *ti-ngal mo porooq.*

3p-do/get dance that 3s-end and 3p-pound taro pudding

‘They finished doing that dance, and they made taro pudding.’

(26) *Tamtamon ti-sov ti-vai mako.*

people 3p-all 3p-marry not

‘Not everyone gets married.’

The third person singular form of this verb, *isov*, also functions as the universal quantifier within noun phrases meaning ‘all’ in expressions like *gabua naol isov* ‘everything’, *saawe isov* ‘all the time’, and *taan isov* ‘all the land’, ‘everywhere’.

(27) *Saawe isov, ti-gham-gham tauvene.*

time all 3p-RDP-do like.that

‘All the time they are acting like that.’

2.1.2 Serial parataxis encoding phases of a single event

In addition to these quasi-grammatical functions of the paratactic serial construction, it is also used in a number of highly lexicalised (and textually very frequent) expressions to indicate sequences of events which commonly co-occur, exhibit high temporal continuity — that is, they occur in close temporal succession or are simultaneous — and which are all more or less under the control or influence of the agent or effector3 in the initial event and proceed towards the same ultimate locative goal or resulting state. Actually, given

3 The term effector here follows Van Valin and LaPolla (1997:118), who characterise this semantic role in the following manner: ‘This labels the participant that brings something about, but there is no implication of its being volitional or the original instigator. It is simply the effecting participant. It need not be animate.’
Pawley’s (1993:108) definition of an event as a ‘bounded happening’, these event serialisations are best viewed as being ‘sub’-events or phases of a single complex event. The sub-events exhibit extremely high action continuity/coherence, and the boundaries of the complex event as a whole occur at places of transition, where there are changes to different actors, locations, directions, times, etc. Unlike idioms, these highly lexicalised sequences of events seem to be semantically very transparent.

According to Gestalt theory, multiple entities exhibiting good continuity are more likely to be perceived as a single entity. Croft (2001:337) gives the following visual example.

![Visual example of continuity](image)

Because the two line segments in the figure on the left are aligned in the same direction, they exhibit high continuity and are much more likely to be perceived as constituting a single line than the two segments in the figure on the right, even though the line on the right could actually have been drawn as a contiguous figure underneath the circle:

![Visual comparison of continuity](image)

To delineate the semantics of different encodings of multiple events in Tuam, a simple binary distinction between events (=successive happenings with lower continuity) and sub-events (=successive happenings with high continuity) is not adequate, however. It seems necessary to postulate multiple degrees of continuity. Any kind of linguistic conjunction implies at least some sort of continuity or shared properties between the conjuncts. Wierzbicka (1980:229), in her article on coordination and conjunction reduction, approvingly cites the following comment by the medieval semanticist Peter of Spain regarding the sentences: Sortes and Plato are white, and Sortes is white and Plato is white:

When I say: ‘Sortes and Plato are white’ the conjunction ‘and’ affirms their simultaneity and oneness. And on that account it unites them in whiteness.


In Peter of Spain’s view, a phrase like “Sortes and Plato” (in “Sortes and Plato are running.”) designated one unit composed of two parts ... In other words, the sentence “Sortes and Plato are running” (“Sortes and Plato are white”) is about the unit composed of Sortes and Plato, the sentence “Sortes is running and Plato is running” (“Sortes is white and Plato is white”) is not about the unit composed of Sortes and Plato.

According to Wierzbicka, conjoined elements must have some sort of semantic ‘common denominator’ which allows them to be conjoined into a unit. Specifically, it must be possible to plausibly view them as being parts of a single entity. But notions like ‘common denominator’ or ‘part of’ are relative. A finger is a part of the body, but it is even more a part of the hand. Cognitive linguistics makes use of the notion of scopes of
A notion like PART OF applied to the domain of the body exhibits the following scopes:

knuckle = PART OF finger = PART OF hand = PART OF arm = PART OF body.

An example like A body has ten fingers sounds odd compared to A body has two arms, because finger is outside the normal scope of the notion PART OF BODY. Similarly, long complex sequences of events are made up of smaller and smaller sequences of events and sub-events, with ever increasing degrees of continuity. Paratactic serialisation is used to express sequences of sub-events with the greatest common denominator, that is, the highest amount of continuity.

One extremely common class of paratactic serialisation involves a motion verb that is followed by another verb with the same subject expressing an event with the same actor that takes place in close temporal proximity to the event expressed by the motion verb.

(28) Yes ti-la ti-gham uraat i-la uum.
3p.NOM 3p-go 3p-do work 3s-go garden
‘They went [and] did the work in the garden.’

In example (28) uniform motion towards a single locational goal is expressed: the garden. The motion is for the purpose of engaging in the action expressed by the following serialised clause, and this purpose is in fact realised. If the purpose is not yet realised, the verb -ghe ‘say, think, want, be about to’ is interposed under the same intonational contour with no intervening pause, as in the following example.

(29) Yes ti-la ti-ghe ti-gham uraat i-la uum.
3p.NOM 3p-go 3p-say/want/think 3p-do work 3s-go garden
‘They went [intending] to do the work in the garden.’

It is not a requirement, however, that all the verbs in serialisation constructions have the same subject. It is also possible to have a ‘pivot’ construction in which an agent does something to a theme or addressee and the following verb or verbs express consequent events or resulting states in which the theme or addressee participates. In a localistic sense, such event sequences can be viewed as a kind of journey whose starting point or SOURCE is the causing event and whose end or GOAL is the final resulting state of the affected theme/addressee/benefactive participant. In the next example, the event of hanging up a lamp is broken up into three subcomponents: (1) initial action, (2) resulting path of motion of the theme, (3) ultimate resulting state.

(30) Ti-saron lam i-za-la i-tuuk saanya.
3p-put.up lamp 3s-ascend-go 3s-hang high
‘They hung the lamp up high.’

1. The inceptive sub-event is encoded with the verb -saron,4 which expresses that an agent acts and begins moving something in his or her hands in a direction upwards from his or her head.

2. This results in the lamp moving along an upwards path. The motion verb -zaa indicates this path.

---

4 There are two other verbs in the language expressing the notion of lifting: ndae ‘lift something up off the ground’ and it ‘lift (generic)’.
3. The ultimate result of the action of the agent is that the lamp is in a new state of hanging from some high place. This resulting state is given by the stative verb -tuuk ‘be hanging’.

Although the second and third verbs have different subjects from the first verb, the agent of the first event is still very clearly in control of the events expressed by these second and third verbs. And all three verbs express events proceeding along a monodirectional path of motion to the same ultimate target: the place where the lamp will be hung.

Another very common paratactic serialisation is -rav...-maat ‘hit...die’, the only means in the language for expressing the notion of ‘kill’. Here, the two verbs express a causing action impinging on a theme, which results in an immediate change of state.

(31) I i-rav zi ti-maat gha siŋ toozi i-ma-liŋ.
3s.NOM 3s-hit 3p.ACC 3p-die SER blood 3p.LOC 3s-DTR-spill
‘He killed them and their blood poured out.’

This paratactic serialised construction for ‘kill’ is especially favoured when the person or thing being killed is a highly activated referent that is referred to with a pronoun or just the transitive suffix -u on the verb -rav as in the following example:

3s-live PRH 2p-hit-TR 3s-die
‘He must not live. Kill him.’

If, however, the object referent is less activated/more discontinuous (and therefore expressed by a full noun phrase (33)), or additional linguistic material occurs that is associated with the first verb (34), then normally the conjunction gha is added between the two clauses.

(33) I-rav taazi marani gha i-maat.
3s-hit younger.sibling:3s.GEN small SER 3s-die
‘He killed his little brother.’

(34) Ti-rav-u pa maet gha i-maat.
3p-hit-TR OBL stone SER 3s-die
‘They stoned him to death.’

In the paratactic serial construction, it is not necessary for the participant inciting the event sequence to have absolute control. He/she/it only needs to have primary control.

In the next example, the stick standing in the water clearly has no volition. It exhibits an effector semantic role. But it does exhibit a strong restraining influence which prevents the canoe from moving.

(35) Ti-ngun ngun pa waan. Leso i-kis waan i-pot
3p-stick vertical.stick OBL canoe so.that 3s-hold canoe 3s-float
be.firm EMPH.LOC over.there
‘They stuck a stick vertically [in the water] for the canoe to keep the canoe floating securely in that place over there.’
This example is also important, because here the combined verbs in the serialisation construction encode events or situations that are not temporally sequenced. The holding of the stick, floating of the canoe, and its being securely in one place are all more or less simultaneous. Still, there is a causal/resultative relationship between the holding and the floating securely in one place.

Summarising what we have observed for these uses of paratactic serialisation, we can say that this construction is used in syntactically regular, semantically transparent formulaic expressions indicating a highly continuous series of subevents that make up a single complex event, namely situations where: (1) there is very little temporal gap, if any between the events, (2) there is some sort of causal/reason-result relationship among the events, (3) the key participants in the event sequence are all present in the first clause, (4) the agent or effector in the initial sub-event has very pronounced control or influence over the whole chain of sub-events, and (5) the sub-events take place more or less in the same direction.

2.2 Conjunction of clauses with gha

The next type of event encoding is that which utilises the conjunction gha. In many ways, this construction seems to overlap functionally with serial parataxis, since it also occurs in constructions encoding highly continuous combinations of events, and the combination normally occurs under a single intonational contour with no intervening pause between the conjoined units. Thus there are often two possibilities for saying something, one of which lacks the gha, while the other has it. For example, in the following sentence, it is possible to omit the gha.

(36) Ti-ziir zi (gha) ti-ghau ti-la.
   3p-send.away 3p.ACC SER 3p-go.away 3p-go
   ‘They sent/drove them away.’

In other instances, however, like the following one, gha would be obligatory.

(37) Ti-gham le aniiŋ gha i-ghan,
   3p-do/get RECX:3s.GEN food SER 3s-eat
   ve tapiri i-vot muul.
   and strength:3s.GEN 3s-appear again
   ‘They gave him [some] food to eat, and his strength returned.’

Contrast the following two examples, which could both be used to describe the same real world event of parents going to get their child and bringing him or her home.

(38) Ti-gham natu-zi i-muul i-lam.
   3p-do/get child-3p.GEN 3s-return 3s-come
   ‘They brought their child back.’

(39) Ti-gham natu-zi gha yesŋa ti-muul ti-lam.
   3p-do/get child-3p.GEN SER together.they 3p-return 3p-come
   ‘They got their child and together they came back.’

In (38), the child is being portrayed as a more or less passive participant, whose returning is completely under the control of his or her parents. In (39) however, he or she is being portrayed as a co-operating participant in the group that acts together. Gha is
extremely common when the participants associated with the first verb are combined into a
group which then functions as the subject of the second verb. The combination of separate
participants into a group seems to be treated as a different participant. The introduction of
this different participant introduces a slight discontinuity into the event sequence and
therefore requires the use of *gha*, even though action and temporal continuity is
maintained.

Both paratactic serialisation and conjunction with *gha* are used to express what seem to be
phases of a single complex event that is mainly under the control or influence of the
subject of the first verb. The causative relationship between the two events expressed by
both constructions seems highly expected. The second event is a commonly occurring
outcome of the kind of event expressed by the first clause. But with *gha*, there normally
seems to be an indication of slightly less control on the part of the subject of the first verb
over the situation expressed by the second verb, as in (39) above. But this is not invariably
the case, as the following extremely common example illustrates:

(40)  *I-gham aniŋ gha i-ghan.*

3s-do/get food SER 3s-eat

‘He took the food and ate it.’

Here the subject of the first clause is obviously in complete control of both the taking
and the eating of the food. The only explanation I currently have for the use of *gha* here is
that *A gham B ghan* without the *gha* would mean that A somehow made/forced B to eat. In
an example like (40), this would mean that the person made the food to eat, which would
be nonsense. In this example, it would also be possible to substitute the general co-
ordinating conjunction *ve* ‘and/but’ for the *gha* with no significant change in meaning.

As was noted in the preceding section, there is a formal factor that favours the use of
paratactic serialisation: it is preferred when the preceding verb phrase is minimal, usually
because the referent of the object is high activated in the current linguistic context. This,
too, is a kind of continuity: of referents. The more linguistic material there is in that verb
phrase, the more likely it is that *gha* will be used, particularly when the object is a full,
non-pronominal noun phrase (see examples (33) and (34) above).

The examples of conjunction with *gha* thus far have all been of clauses encoding event
sequences. But like the paratactic serialisation construction, *gha* can also encode
simultaneous events, as long as there is some sort of causal link between the two, often one
of result.

(41)  *Ti-laagh taghon eez gha ti-la.*

3p-walk follow road SER 3p-go

‘They walked further along the road.’

And like the paratactic serialisation construction, clauses conjoined with *gha* exhibit
obligatory sharing of operators.

(42)  *Yes i-rau ti-mat-maat gha ti-la le-z mako.*

3p.NOM 3s-able 3p-RDP-die SER 3p-go RECX-3p.GEN not

‘They are not going to die and perish.’

In my text corpus, the most common pairs of verbs/predicates conjoined with *gha* are:
<table>
<thead>
<tr>
<th>Verb 1</th>
<th>Meaning</th>
<th>Verb 2</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ngarig</td>
<td>‘now’</td>
<td>la</td>
<td>‘go right’</td>
</tr>
<tr>
<td>gham</td>
<td>‘do’/‘get’/‘cause’</td>
<td>motion verb</td>
<td></td>
</tr>
<tr>
<td>burig</td>
<td>‘get up’</td>
<td>motion verb</td>
<td></td>
</tr>
<tr>
<td>taghon</td>
<td>‘follow’</td>
<td>motion verb</td>
<td>‘follow someone’</td>
</tr>
<tr>
<td>rav</td>
<td>‘hit’</td>
<td>maat</td>
<td>‘die’</td>
</tr>
<tr>
<td>gham</td>
<td>‘do’/‘get’/‘cause’</td>
<td>ghan</td>
<td>‘eat’/‘consume’</td>
</tr>
<tr>
<td>savia</td>
<td>‘speak about’</td>
<td>loon</td>
<td>‘listen’/‘hear’</td>
</tr>
<tr>
<td>pul</td>
<td>‘leave’/‘let’/‘allow’</td>
<td>motion verb</td>
<td></td>
</tr>
<tr>
<td>burig</td>
<td>‘get up’</td>
<td>vazorai</td>
<td>‘dispute with each other’</td>
</tr>
<tr>
<td>gham</td>
<td>‘die’</td>
<td>la le</td>
<td>‘go away’/‘be lost’/‘perish’</td>
</tr>
<tr>
<td>gham</td>
<td>‘do’/‘get’/‘cause’</td>
<td>muul</td>
<td>return</td>
</tr>
<tr>
<td>vazorai</td>
<td>‘dispute with each other’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gham</td>
<td>‘die’</td>
<td>la le</td>
<td>‘go away’/‘be lost’/‘perish’</td>
</tr>
<tr>
<td>saav</td>
<td>‘collect’</td>
<td>gham</td>
<td>‘die’</td>
</tr>
<tr>
<td>saav</td>
<td>‘speak’</td>
<td>yau</td>
<td>‘collect’</td>
</tr>
<tr>
<td>dit</td>
<td>‘go out’</td>
<td>ghita</td>
<td>‘see’</td>
</tr>
<tr>
<td>saav</td>
<td>‘speak’</td>
<td>mbole izi</td>
<td>‘sit down’ (lit. ‘arse descend’)</td>
</tr>
<tr>
<td>reu</td>
<td>‘destroy’</td>
<td>la le</td>
<td>‘go away’/‘be lost’/‘perish’</td>
</tr>
</tbody>
</table>

One final note regarding gha is that it also functions as an alternative phrasal co-ordinator within noun phrases to the co-ordinating conjunction ve ‘and’. Gha is used to conjoin a pair of commonly associated participants, for example the names of a pair of brothers, or the names of a husband and a wife, or the nouns tama ‘father’ and tina ‘mother’.

2.3 The telic conjunction le

The next connective in narratives to be considered is the telic conjunction le. It always indicates chronological sequence and is used to express that the end point of an event sequence has been reached. The end point may be temporal as in the following example.

---

5 The conjunction le should not be confused with the inalienable noun le-, whose third person singular form is also le. The inalienable noun inflects: leg, lem, le, leen, lemai, lelim, lezi(i) and is used to: (1) encode recipients, (2) assert, deny, or question possession, (3) express that a participant undergoes some sort of pleasurable experience, and (4) in the construction -la le- ‘go RECX’ indicate that a participant somehow goes away or perishes against their will.
Event integrity and continuity in the Tuam language

(43) Ti-laagh ti-la le mboŋ i-zi.
3p-walk 3p-go TEL night 3s-descend
‘They walked along until night came.’

The end point may also be aspectual. The very common sequence le + -sov ‘end, finish, all’ expresses completive aspect.

(44) I i-sav-saav le i-sov, ve eval tiina to lup-u-uŋ
3s.NOM 3s-RDP-speak TEL 3s-end and group big.one LOC meet-TR-NR

\textit{tana ti-neneey mon.}
that 3p-be.quiet only
‘After he finished speaking, the crowd in that meeting was completely quiet.’
[i.e. They were unable to say anything in response.]

\textit{Le} is also used to encode degree or extent of action.

(45) Yes, raa, pitool zi le ti-saghat kat.
3p.NOM poor hungry 3p.ACC TEL 3p-be.bad very
‘They, poor people, were really hungry and in a bad way.’

In accordance with this function of expressing the extent or degree of an action, the construction -ghe le VP, is used to express that someone wants to do something very much/to a great extent.

(46) Yau na-pasawal na-ghe le na-ghit gham, o poia.
1s.NOM 1s-long 1s-want TEL 1s-see 2p.ACC then good
‘I am really longing to see you (PL). If only I could see you then [it would be] good.’

Finally, the end point may be an event, usually a result of or response to the event(s) expressed by the clause(s) preceding the \textit{le}.

(47) I i-sav-i-a sav-e-η tana
3s.NOM 3s-speak-APP-TR speak-SV-NR that

\textit{le i-reu lolo-zi ve oroɔ tiina i-burig.}
TEL 3s-destroy insides-3p.GEN and uproar big.one 3s-get.up
‘When he said that, it really upset them and a big uproar began.’

Since \textit{gha} can also express result, the question arises, what is the difference between results expressed by \textit{gha} and those expressed by \textit{le}. In some instances, the two conjunctions may be used interchangeably, but \textit{le} is less frequent in texts than \textit{gha}. This difference in frequency suggests that \textit{le} has more semantic components than \textit{gha}, which restrict its collocational possibilities. Some of these additional components are aspectual in nature: duration and completion. \textit{Le} is favoured following clauses encoding situations or events that are more durative in nature and which have a clearer end point, whereas the resulting events expressed by clauses conjoined with \textit{gha} typically follow immediately from more punctiliar actions and may not be as clearly bounded. Also, whereas in paratactic serialisation and conjunction with \textit{gha}, the result is always \textit{intended} by the agent of the first clause, with \textit{le}, the result may or may not be an intended one. Thus, in example (47) above, inciting a riot was not the intention of the speaker. It was an \textit{unintended} result/effect of what he said. In the following example, the notion of intention is clearly absent from both the reason and the result clauses.
(48) Morogh-o-oŋ tiina i-gham-u le i-gharau i-ghe i-maat.
be.sick-SV-NR big.one 3s-do/get-TR TEL 3s-approach 3s-about.to 3s-die
‘A severe illness afflicted him with the result that he was on the verge of dying.’

End points introduced by le may also be geographic in nature. Thus, stopping points or the conclusion of a series of events happening in the same direction are often encoded with le. In the following example, note that the clause expressing the end of the ascent phase of the journey up the mountain is conjoined with le. The descent phase, a new direction, is introduced with the conjunction ra, which expresses discontinuity/the beginning of a new event sequence.

(49) Ti-laagh ti-la le ti-za-la loloz daaba.
3p-walk 3p-go TEL 3p-ascend-go mountain head:3s.GEN
ra ti-ndari ti-zi-la.
then 3p-go.down.at.an.angle 3p-descend-go
‘They walked up to the top of the mountain, and then they started descending.’

Frustration, the fact that an event sequence did not lead to the anticipated result, is expressed by the sequence le mako ‘and not’.

(50) Ti-il aniiŋ le mako.
3p-search.for food TEL not
‘They searched for food without finding any.’ /‘They searched for food in vain.’

Note in this example that the scope of the negative mako does not extend all the way back to the predication ‘they searched for food’. The following example shows a similar limitation of negation scope.

(51) Ti-looŋ sav-e-eŋ tooni le i-rau lolo-zi mako.
3p-hear speak-SV-NR 3s.LOC TEL 3s-sufficient insides-3p.GEN not
‘They heard what he said, but it was not acceptable to them.’

Thus, conjunction with le does not exhibit the same sort of operator dependence as the paratactic serialisation construction and conjunction with gha, even though clauses conjoined with le still seem to express phases of relatively continuous event sequences, specifically the final part of the sequence.

In extended narratives, le frequently concludes the events of the setting section of a discourse and marks the transition to an inciting event of the story. Often the clause following the le- is a presentative clause containing the predicates -zaa ‘ascend to’/‘come upon’, or -ghita ‘see’, which introduces a further key participant in the narrative.

(52) Ti-yaat ti-zi pa waanŋ, ve ti-laagh ti-la
3p-jump 3p-descend OBL canoe and 3p-walk 3p-go
le ti-zaa to yeer eez.
TEL 3p-ascend LOC man one
‘They jumped down from the canoe, and walked a bit, and came upon a man.’

(53) Na-gud waanŋ i-zi-la te, ve na-looŋ na-la
1s-drag canoe 3s-descend-go sea and 1s-enter 1s-go
‘I dragged the canoe down to the ocean and went in towards
pa Dgure. Na-la le na-ghit-a iigh ti-rav mban.
OBL Nggure 1s-go TEL 1s-see-TR fish 3p-hit bait
the Nggure river. As I went, I saw that fish were driving small bait-sized fish

ve ti-loog pa taan.
and 3p-enter OBL land
and coming in towards the shore.’

Following this example, the author goes on to tell how he went in to the shore, got his fishing equipment, and went out to get the bait fish. Then he sees a much bigger fish and spears it. But because he mistakenly identifies it, he is not sufficiently careful in grabbing it, and ends up getting badly cut by the fish’s fins. This example is typical in that the transition from the end of the setting of the narrative to the inciting event of seeing the bait fish is marked with _le_.

### 2.4 Co-ordination with the conjunction _ve_ and co-ordinate parataxis

The most semantically unmarked conjunction in the language is _ve_ ‘and’. By way of comparison, in texts it is roughly six times more frequent than the next most common conjunction _gha_. Its great frequency testifies to its relative lack of semantic specificity. It frequently co-ordinates clauses expressing sequential events as in (54).

(54) _I-vot i-la, ve i-gherev puugh toman iigh-a_
3s-go.out 3s-go and 3s-pull net with fish-ASS:3s.GEN
_i-loog i-lam taan._
3s-enter 3s-come land
‘He went out [into the water], _and_ pulled the net with the fish in it, back to the shore.’

Simultaneity plus emphasised imperfective aspect expressing progressive, on-going action is indicated via conjunction with _ve_ plus reduplication of the verb as in (55).

(55) _Yesuru ti-savsaav, ve tina-zi taliiña_
3d 3p-speak:RDP and mother-3p.GEN ear:3s.GEN
_i-saut sav-e-en toozi._
3s-catch speak-SV-NR 3p.LOC
‘As the two of them were speaking, their mother overheard their talk.’

_Ve_ is also used to conjoin atemporal, gnomic clauses.

(56) _Lasom, ene tintiina, ve nimmim ne geegeu._
blue.fly this.one big:RDP and fly this.one small:PL
‘Blue flies [are] big, _and_ ordinary flies, these [are] small.’

Compared with the conjunctive devices discussed in preceding sections, _ve_ is used to conjoin clauses expressing events having relatively less continuity. Thus, when the second clause exhibits a change of subject and the new subject’s referent is not mentioned in the preceding clause, _ve_ is typically used.

(57) _Yei ni-raav, ve i i-laagh taan._
1pe.NOM 1pe-sail and 3s.NOM 3s-walk land
‘We (EX) sailed, _and he_ walked on the land [to the same destination].’
If the same subject is maintained across clauses conjoined with ve, then the action expressed is somehow less continuous. In such instances, instead of conjoining different phases or parts of a single complex event, ve marks the beginning of a new event sequence, one which is not simply a further continuation of the current event sequence. There may not be a clear causal relationship between the events of the two clauses, or a different direction of motion may be taken as in the next example.

(58) I i-wai i-la i-vot maghat paavo,
3s.NOM 3s-swim 3s-go 3s-go.out shallow.spot surface:3s.GEN
‘She swam out to a shallow spot in the ocean [where one can walk on the surface of the reef],

ve i-zeuzeu taghon zeet-ŋa zi-livaa siriv
and 3s-forage.on.reef follow associate:3s.GEN-PL PL-female some
‘and [=transition from swimming event to foraging event] foraged on the reef following some of her girlfriends
tau ti-zeuzeu gha ti-muŋ we
that 3p-forage.on.reef SER 3p-precede REM.PRG
‘that were foraging ahead of her off in the distance.’
[Since the island is surrounded by a reef, one goes out to the reef and then makes a right angle turn to follow along the edge of the reef.]

Example (59) comes after a sentence stating the death of the father in the presence of the son. The son’s seeing the death of his father in the first clause is therefore presupposed.

(59) I i-ghit-a taama i-maat, ve i-tap i-za-la
3s.NOM 3s-see-TR father:3s.GEN 3s-die and 3s-fall 3s-ascend-go

paavo, Ø i-saghav-u, ve i-taŋ.
surface:3s.GEN 3s-embrace-TR and 3s-cry
‘[When] he saw [that] his father had died, he fell on top of him, Ø embraced him, and cried.’

Example (59) comes after a sentence stating the death of the father in the presence of the son. The son’s seeing the death of his father in the first clause is therefore presupposed. It is common to conjoin such initial presupposed clauses (often they are heads in tail-head linkages, which repeat the final clause of the previous sentence) with ve. This construction almost seems to function like an alternate sort of temporal subordinate clause.6 Perhaps the use of ve here is due to the fact that such clauses do not really advance the event line of the narrative. Ve, then, could be viewed as marking the transition from the presupposed material back to the assertions of the event line of the narrative.

Clauses conjoined with ve do not need to have the same participants (see (56), (57) above), nor do they exhibit operator dependence for negation or modality (60).

---

6 The other possibility for temporal subordinate clauses involves relative clauses that modify the noun saawe ‘time/space/lifespan’.
Event integrity and continuity in the Tuam language

(60) Ti-ghur muri-zi pani, ve ti-uul-e mako.
3p-put back-3p.GEN 3s.OBL and 3p-help-TR not
‘They gave their backs to him [i.e. rejected him], and did not help him.’

Whereas gha conjoins pairs of noun phrases whose referents are closely associated, ve has no such restriction when used as a co-ordinator within the noun phrase. It can conjoin more than two phrases, and the referents of the phrases are not as closely associated as in the case with gha. Still, as Wierzbicka (1980) notes, they must share at least some common property or properties in order for their co-ordination to be felicitous.

There are two quasi-grammatical uses of ve which seem to contravene its normal usage of encoding distinct events. In both of these constructions, it alternates with gha: (1) -leep ‘stay, live, exist’ ve/gha S is the only means for expressing locative source with motion verbs (see (61)) and (2) gham ‘do/get’ A ve/gha S functions as an alternative way of encoding instruments (see (62)).

(61) Ra naatu olman aa-na i-leep uum ve i-lam.
then child:3s.GEN older turn-3s.GEN 3s-stay garden and 3s-come
‘Then it was his older son’s turn to come from the garden.’

(62) Ti-leep yaa loolo gha ti-zaa.
3p-stay water inside:3s.GEN SER 3p-ascend.
‘They came up out of the water.’

(63) Ti-gham maet tana, ve ti-raav mbun toozi.
3p-do/get stone/money that and 3p-hit debt 3p.LOC
‘They used that money to pay off their debt.’/‘They paid off their debt with that money.’

(64) Ti-gham waan toozi gha ti-raay ti-la pa sisi ite.
3p-do/get canoe 3p.LOC SER 3p-sail 3p-go OBL island other
‘They took their canoe and sailed to another island.’

Perhaps the reason that ve is used in the locative source construction is that the clause with -leep is not, strictly speaking, an event. It is the starting point/state from which an event chain begins. At present, I have no explanation for the use of ve as an alternative conjunction in the instrumental construction.

2.5 The conjunction ra(gha) ‘(and) then’

The conjunction expressing the greatest amount of discontinuity between events is ra(gha) ‘then’. (The abbreviated form, ra, is more common in texts than the full form.) It never conjoins clauses expressing simultaneous events and always expresses temporal sequence. Ra is favoured when: (1) there is a change of subject, (2) there is a significant temporal gap (in which case it often occurs in the phrase murei ra ‘sometime afterwards’ (lit. ‘later, then’)), and/or (3) a new sequence of events directed towards a different location begins. In many instances, however, ve and ra(gha) seem to be freely substitutable.

---

7 The more common way of expressing instruments is via oblique prepositional phrases: pa maet ‘with stone/money’.
Naviu malmali Ø i-za-la pa nughei,  
Naviu directness:3s.GEN 3s-ascend-go OBL village  
‘Naviu went directly up to [her] village’

ra tiina i-seŋ ygool aavo,  
then mother:3s.GEN 3s-hit shellfish mouth:3s.GEN  
‘and then her mother struck the mouth of the shellfish [which had clamped onto to Naviu’s hand],

ve i-pas niima pani gha i-vot.  
and 3s-pull.out hand:3s.GEN 3s.OBL SER 3s-go.out  
‘and pulled her hand out from it.’

(Note here that in this one sentence expressing a succession of events occurring in relatively close temporal sequence, four different connective devices are used: Ø (=juxtaposed serialisation), ra, ve, and gha.

Ti-gham sav-eŋ pa-roiŋ pazi muul,  
3p-do/get speak-SV-NR CAUS-fear-NR 3p.OBL again  
ra ti-pul zi ti-la.  
then 3p-let 3p ACC 3p-go  
‘They threatened them again, and then they let them go.’

3 Summary of the distinctions between different connective devices

By way of summary, the principle distinctions between the various connective devices are tabulated below.

<table>
<thead>
<tr>
<th>Device</th>
<th>Can express close temporal sequence</th>
<th>Can express simultaneity</th>
<th>Degree of control of first clause subject</th>
<th>Operator dependence (negation, modality)</th>
<th>Expresses causality</th>
<th>Goal of motion/action</th>
<th>Obligatory sharing of participants</th>
<th>Can express larger temporal gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>parataxis</td>
<td>+</td>
<td>+</td>
<td>complete</td>
<td>+</td>
<td>+</td>
<td>same</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>conjunction with gha</td>
<td>+</td>
<td>+</td>
<td>high</td>
<td>+</td>
<td>+</td>
<td>similar</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>conjunction with le</td>
<td>+</td>
<td>-</td>
<td>variable</td>
<td>-</td>
<td>+</td>
<td>similar</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>co-ordinate parataxis</td>
<td>+</td>
<td>+</td>
<td>variable</td>
<td>-</td>
<td>-</td>
<td>n/a</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>conjunction with ve</td>
<td>+</td>
<td>+</td>
<td>variable</td>
<td>-</td>
<td>-</td>
<td>n/a</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>conjunction with ra(gha)</td>
<td>+</td>
<td>-</td>
<td>variable</td>
<td>-</td>
<td>-</td>
<td>n/a</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
The implication of this table is that there is a cline of event continuity in the language.

<table>
<thead>
<tr>
<th>Most continuous</th>
<th>Ø (serial parataxis)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>conjunction with gha</td>
</tr>
<tr>
<td></td>
<td>conjunction with le</td>
</tr>
<tr>
<td>Co-ordinate parataxis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>conjunction with ve</td>
</tr>
<tr>
<td>Least continuous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>conjunction with ra(gha)</td>
</tr>
</tbody>
</table>

One important component of the skill of ‘saying things’ idiomatically (Grace 1981) in Tuam is being able to make judgements of event continuity. The connective system of the language forces its speakers to constantly consider the relative continuity of events. In addition, use of the telic conjunction le requires them to consider the boundaries of event chains, since it indicates the conclusion of an event sequence. In the case of paratactic serialisation, most instances of this construction seem to be highly lexicalised units, with little possibility of adding an intervening conjunction between the clauses. In the case of the quasi-grammatical instances of this construction, there is no possibility whatsoever of adding a conjunction. But with the conjunctions gha, le, ve, and ra(gha), the speaker has more freedom. With this freedom comes the possibility of disagreement on the part of native speakers over what conjunction would be best in a given context. Still, there are limits. Pairs of connective devices that are adjacent to each other in the above chart can occasionally be used interchangeably, but those separated by some other device seem to be very clearly differentiated. Thus gha can sometimes alternate with le, le can alternate with ve, and ve can alternate with ra(gha), but ra(gha) would never alternate with gha.

I will conclude with a final example from the Tuam translation of the gospel of Luke which illustrates well the use of most of these connective devices.

(67) Yesu i-saav tauvene, ve ti-kis-i gha
Jesus 3s-speak like.this and 3p-hold-TR SER
‘Jesus spoke like that, and [beginning of new event sequence having to do with Jesus’ arrest] they took hold of him, and [beginning of the transport sub-phase of the arrest sequence]

ti-gham-u Ø ti-la le ti-looŋ Ø ti-la
3p-do/get-TR 3p-go TEL 3p-enter 3p-go
‘took him [path] away and [end of the transport phase and of the arrest sequence as a whole] entered into
daaba to watoonŋrau ruum tooni.
head:3s.GEN LOC offering house 3s.LOC
‘the house of the chief priest.’
Schematically, the perspective on events portrayed in this sentence can be represented as follows:

References


Aiyakoe beká resena! [Korafe-Yegha] Aïye bee rere! [Tafota Baruga] ‘Thank you truly I am saying’, Andy! You were always challenging me to search deeper and find the gold. You usually found my ‘gold’ somewhere in the midst of the morass of a chapter I was sunk in, and identified the starting point for the chapter. A well-rounded individual who sees linguistic possibilities in every area of life (cricket announcing, tourist brochures, igo igo, …), you encouraged us also to think outside the box. Inspired to look for weird English constructions like you were continually discovering, I found a strange tense juxtaposition in a letter a friend wrote:

I’m still looking for another job. As a matter of fact, tomorrow I took the day off work and am going to let my fingers do the walking as I make dozens of phone calls to inquire of openings.

(The only interpretation I can give for this is that the writer already has taken the day off with her supervisor, but she has not yet actually begun her day off at the time of her speech communication act.)

It wasn’t all work and no play either. You could take my husband, Jim, apart on the tennis court. We both are indebted to you for teaching us ‘Americans’ the game of cricket.

You opened up Pandora’s box with unconventional conversation stoppers like ‘The clause¹ is a convenient fiction.’ That may have been said tongue in cheek, but I have pondered it since as I have tried to identify clause boundaries in serial verb constructions and in grammatical constructions composed of at least one medial verb with an auxiliary verb.

Contiguous serial verbs in a string tend to yield a ‘well-behaved clause’ for Korafe-Yegha,² but the Kalam serial strings you cited are more challenging than the Korafe ones! Is the following verb string with nine verb stems and a final verb one clause or something more than a clause?

¹ Pawley (1993:120) gives a more conventional statement about the clause: ‘The clause is a very useful analytic construct, but we should take care not to reify it or to force all clause-like constructions into its mold. There is a cluster of properties that can be associated with typical clauses. In any one language, some constructions will have more of these diagnostic or typical properties, some only a few, and others none.’

² Yegha is currently considered a dialect of Korafe. Henceforth, I will just use Korafe.
Korafe averages four or less verbs in a serial string. Example (2) is a Korafe sentence with two serial verb constructions broken up by the motion verb *fera* ‘they came’.

(2) … *taima=da ofo keve bu f-era*

*bush=LOC flooring shave:*I *get:*I *come:*DUR-PST.SEQ.R.3p

*fuge tendud-useri.*

*throw:*I *lash:*II-DP.3p.AQ

‘… in the bush they shaved off the bark of the flooring, brought it, threw it (onto the joists) and lashed it securely.’

The verb *come* is in a past sequencing medial verb form. Korafe only has medial forms, not serial forms, for the verbs *furari* ‘come’ and *yari* ‘go’. Medial verbs are marked to indicate a temporal distinction (either sequencing or overlapping) between the verb in the marking clause and the verb in the referencing clause. Therefore, medial verbs routinely signal a clause boundary. If the Kalam sentence above patterned like Korafe, the verb *ap* ‘come’ would signal clause junctures and distinguish three clauses in example (1). Moving from one venue to another is certainly a possible semantic rationale for establishing clause junctures.

However, Korafe non-contiguous serial constructions are much more problematic than the contiguous serial constructions. How many clauses are in example (3)?

(3) *Fegha viti kambo joká=da bayau okia=da vendi*

*Fegha ascend:*I *house inside=LOC food claypot=LOC put.in:*I

*bu vose avaraka use it-iri …*

*get:*I *descend:*I *fire fan/blow.up:*I *cook:*I-SEQ.R.3s.DS

‘Fegha went up into the house and put food into a claypot, got it and went down, got the fire going and cooked it …’

*Fegha* is the subject for the entire string, but *kambo jokáda* ‘inside the house’, *bayau ‘food’ and okiada ‘in the pot’ are arguments closely related to *vendi* ‘put inside’. The object argument *bayau ‘food’ has the possibility of extending beyond *vendi* ‘put inside’ to *bu ‘get’, but it has little relationship to *vose ‘descend’ or use ‘blow up’ which has its own local object *avaraka ‘fire’. It is understood that *bayau ‘food’ is the object of *itiri ‘he cooked’. The story teller chose to use all verb stems in this construction up until *itiri*, but all of the verbs only share one argument; other arguments are more locally based. So is this one clause or something else?

Or on the other hand, medial verb strings mesh to form one grammatical unit encoding an aspectual iteration of the same event(s) at irregular intervals:
The SS medial suffix -do, arising from do ‘leave off’ signals the intermittent nature of the event sequence. The temporal period within which it is repeated is indicated by whatever tense or medial form is marked on ghe. It is understood that other activities can and do intervene. For instance, in example (4) it is understood that the speaker could walk during parts of his run for exercise.

(4) Na oro ko tamo gase gi+u-se, sum bu-do gh-a re na.  
1s today body exercise+do:1-SIM SS run:1-SEQ SS do.again:1-FUT 1s.FN  
‘While I will be exercising today, I will be running intermittently.’

Two clauses, sumbudu ‘running and then’ and gharena ‘I will do again,’ are used to indicate the grammatical aspectual category, iteration over irregular intervals.

In example (5) the medial verb forms ya ‘go:IRR SS’ and foa ‘come:IRR SS’ are used together with ghuuse ‘while doing again and again’ to encode iteration of the same events at irregular intervals. Although divu ‘dance.1’ appears to be in the serial verb construction bounded by savi-ya ‘went inland’, the use of the ghuuse links buvu-foa ‘came out’ with savi-ya. Also, the action of dancing is understood to occur with buvu-foa as well. Grammatically, there are three clauses linked here to encode iteration over irregular intervals.

(5) 1. nu divu savi+y-a  
3s sing:1 go.inland+go:SEQU SS  
she repeatedly sang and went inland (west) and

2. buvu+fo-a 3. ghu-se, …  
come.out:1+come:SEQU SS do.again:1-SEQ.SS  
came back out (east and while she was doing that),  
‘… while she was repeatedly singing, going inland (west)  
and coming out (east), …’

In the following sentence (6), an entire sequence of 18 verbs (all medial, therefore all clauses in their own right) terminates with the final verb gheteri ‘they repeatedly did’. Semantically, the terminal auxiliary verb gheteri indicates that the set of 18 events it accompanies is repeated in its entirety from time to time.

(6) S-eari, nunda aki-mane vose-do  
say:1-SEQ.CUST.3s.DS 3s.GEN older.sister-PL descend:1-SEQ SS  
‘She would speak and her sisters would come down and

nati+jughu ava  
village+underneath.house that:CT.FOC  
underneath the houses

\[ (X) V^{(0-n)} = \text{verb stem 1-do or } V_{\text{SEQ,CUST,DS}}^{\text{optional}} \]
\[ + (X) V_1=\text{verb stem 1-do}^3 + V_2=\text{ghe ‘do again:1’}. \]

The verbs i ‘go’, fu ‘come’ and iri ‘remain, be’ do not have a non-durative stem that combines with -do ‘leaving off’. The sequencing irrealis SS verb forms ya, fafoa, and iraliria occur instead in combination with ghe ‘do again’.

3
nasara+tafono+eoro
hand.broom+bark.dustpan+do:SEQ.CUST.3p.DS
they would sweep and

dadab-eari.
finish-SEQ.CUST.3s.DS
when that was finished

okia bu-do y-ama
clay.pot get:I-SEQ.SS go:DUR-SEQ.IRR.SS
they would get clay pots and go and

uvu dimbu-do fo-ama fiti-do
water dip.up:I-SEQ.SS come:DUR-SEQ.IRR.SS put:I-SEQ.SS
dip up water and come and put it (at the houses) and

do-do y-ama ika bambu-do
leave:I-SEQ.SS go:DUR-SEQ.IRR.SS wood get:I-SEQ.SS
they would leave and go and fetch firewood and

fo-ama fiti do-do.
come:DUR-SEQ.IRR.SS put:I-SEQ.SS leave:I-SEQ.SS
they would come and put it (at the houses) and leave, and

ade+jamena amo, ne y-ama
girl+PL that:TOP/FUT 3p go:DUR-SEQ.IRR.SS
those girls, they would go and

bijo+Gharube aminda viti-do
banana+Gharube that:CEFF.TOP/FOC.LOC ascend:I-SEQ.SS
they would climb up into the Gharube banana tree

ir-ia+ghe-teri.
remain-IRR.SS+do.again:I-TP.3p.AQ
and they would remain and they kept doing (that).

Not only did you challenge me ‘to think outside of the box’ to accommodate serial and medial verb constructions, the ‘one clause at a time hypothesis’ that you and Syder (1976, 1983:202) proposed for ‘fluent units’⁴ revealed another significant distinction that Korafe exhibits. You described the fluent unit as ‘a technical term to refer to a stretch of pause-free speech uttered at or faster than normal rate of articulation — about five syllables per second in English.’ You credited fluent talkers with pausing or slowing down at the end of each clause of four to ten words. When I took a look at Korafe ‘fluent units’, I discovered that, like English, they also normally have four to ten words in them. Like English units, they also are defined by semantic and phonological criteria. However, in Korafe narrative discourse, they do not necessarily equal a single clause. Sometimes they contain four clauses (i.e. a string of four medial verbs), so I have called them ‘thematic clause chain units’. For instance, example (7) exhibits five thematic clause chain units: unit 1 is a

⁴ Chafe (1985:22) calls these units ‘intonation units’, and Grimes (1975:274) and Litteral (1980:241ff.) call them ‘information units’.
recapitulation of the final verb in the previous sentence and only one clause; unit 2 has four clauses; unit 3 has three clauses; unit 4 two clauses; and unit 5 two clauses.

(7) UNIT 1:
*Kamuse-do*, card.I-SEQ.SS
‘We will card the (inner bark of the *songa* vine) into strips,

UNIT 2:
*bu-do*  *fu-a*,  *iji=da*
get.I-SEQ.SS  come:DUR-SEQ.IRR.SS  sun=LOC
*dend-oro*  *av-ari*,
hang:I-SEQ.R.1p.DS  dry-SEQ.IRR.3s.DS
bring it and hang it in the sun and it will dry, and

UNIT 3:
*ghaito ruve-do*  *fit-oro*
pandanus.strips  coil:I-SEQ.SS  put:I-SEQ.IRR.1p.DS
*ir-ari*,
remain-SIM.IRR.3s.DS
we will coil the (dried) pandanus leaves, put them away (where) they will remain, and

UNIT 4:
*ava*  *bu-do*  *vose*  *besug-edo*,
that:CT.FOC  get:I-SEQ.SS  descend  unfold:I-SEQ.SS
we will get those (pandanus leaves and the *songa* vine) descend and uncoil them, and

UNIT 5:
*tuturo+e-do*  *gemb-arera*.
beginning+do:I-SEQ.SS  weave:FUT.1p.FN
we will begin to weave the mat.’

Thank you for blazing the way for me and others of my colleagues documenting Papuan languages. Your insights have continued to influence me in the current preparation of the Korafe-Yegha dictionary. Since you started studying Kalam and found that it was ‘a language which defies description by ordinary means’ (Pawley 1993), you have searched for and found ‘extraordinary’ means (recording speech formulas [fixed, semi-fixed and productive], determining discourse rules, etc.) to document some of the ‘quintessential’ characteristics of Kalam. I am grateful that you as my supervisor introduced me to your research and enabled me to begin to record the conventions which are part of the idiomatic competence of the Korafe and Yegha peoples, and gave me good ideas for documenting the idiomatic competence of the Tafota Baruga speech community as well.

Now that you will be ‘tidying up’ your office, we await the textbook you are going to write on ‘subject matter lexemes’, recording speech formulas, and formulating discourse structure rules!
References
Pawley, Andrew and Frances Hodgetts Syder, 1976, The one clause at a time hypothesis. (ms presented at the First International Conference of New Zealand Linguistics Society in Auckland, N.Z.)
Wordhood in serial verb constructions: evidence from Anamuxra

ANDREW INGRAM

1 Introduction

The term ‘serial verb (construction)’ has been applied to a range of constructions which vary widely in function and form both within individual languages and cross-linguistically. As Lane (2007:48) observes, serial verb constructions (SVCs) form ‘a heterogeneous class whose boundaries are somewhat ill-defined’. Despite the variety, it is possible to identify a number of ‘core’ criteria which serve as credible diagnostics for SVCs. These criteria are presented together in (1):  

(1) A serial verb construction is a sequence of two or more verbs which act like a single verb in various ways, including the following:

(i) the SVC is a single grammatical unit which forms a single constituent, the predicate, in single verbal clauses.

(ii) each verb can function as the sole predicate in a mono-verbal clause.

(iii) there are no morphemes characteristic of clause boundaries present within the sequence, for example, there is no subordination, co-ordination, or switch reference morphology marking.

(iv) the verbs share at least one argument (typically the subject).

(v) the verbs cannot be separated by intonational cues including pauses normally associated with clause boundaries.

---

1 I am much indebted to the Anamuxra-speaking community for their hospitality. Especial thanks to my main teachers Gabriel Waia, Peter Swakai, Anton Swakai, Thomas Karukai and Patrick Gumagava. My sincere gratitude to Alexandra Aikhenvald, R.M.W. Dixon, Randy LaPolla and anonymous reviewers for extremely useful feedback at various stages of the writing of this paper.

2 The characterisation of SVCs given here draws directly on a number of sources, the most significant of these being: Bradshaw (1982); Foley and Olsen (1985); Crowley (1987); Durie (1997); Lane (1991); Aikhenvald (1999); and Pawley and Lane (1998).

3 A known exception to this is found in ambient serialisation where no arguments are shared (see Crowley 1987).
(vi) the component verbs of an SVC share tense, aspect, mood and often polarity values.

(vii) the SVC may share a single set of peripheral (non-core) adjuncts such as locationals, temporals and so on.

Serial verbs are identified by the clustering of the features listed above. Prototypical SVCs are assumed to display all these properties; less prototypical SVCs may fail on one or more counts, though not with regard to ‘i’ or ‘ii’ (cf. Aikhenvald 1999:470).

Various classifications of SVCs have been put forward based on one or more different parameters. One such parameter is ‘symmetricality’ (Durie 1997; Aikhenvald 1999). According to this feature, SVCs are either symmetrical, that is, each verb in the series comes from an open class of verbs; or asymmetrical, that is, one of the verbs is taken from a restricted subclass of verb types, while the other is taken from an open class of verbs.

Two other prominent and commonly utilised parameters are contiguity and incorporation. Both of these are concerned with the surface configurational possibilities of verb serialisation. Contiguity concerns the sequencing of constituent verbs: whether argument NPs can intervene between the members of the verb sequence or not. Incorporation, on the other hand, concerns the wordhood of the constituent verbs. For Durie (1997) incorporation in wordhood in SVCs is distinctly phonological: that is, the constituent verbs either do or do not form a single phonological word. For others (see Aikhenvald 1999) incorporation involves alignment between phonological and grammatical words such that the contrast lies between SVCs that constitute a single phonological and grammatical word, on one hand, and SVCs whose components are separate phonological and grammatical words, on the other hand.

However, both of these approaches to wordhood in relation to SVCs stand in contrast to the more complex notion of wordhood found beyond SVCs in numerous languages. As argued by Aikhenvald (2006:1):

one should sensibly keep apart phonological criteria, which define ‘phonological’ word and grammatical criteria, which define ‘grammatical word’. In some languages the two types of word coincide and one can felicitably talk of a single unit word, which has a place both in the hierarchy of phonological units and in the hierarchy of grammatical units. In other languages phonological word and grammatical word generally coincide but do not always do so. We may have a grammatical word consisting of a whole number of phonological words, or a phonological word consisting of a whole number of grammatical words. Or there can be a more complex correspondence between the two types of word, with, say, a grammatical word consisting of all of one and part of another phonological word.

---

4 This characteristic is sometimes presented as if it were a ‘problem’ special to serial verbs. However, it is not unique to them. A number of other categories also show this, for example the notions of ‘subject’, ‘subordination’ and so on.

5 A few of the individual descriptions of languages which employ the parameters of incorporation and/or contiguity to classify serial verbs include: Donohue (1999) for Tukang Besi; Hyslop (2001) for Ambae; and Ingram (2001) for Anamuxra.

6 Dixon and Aikhenvald (2002) present a number of cases illustrating complexity in the correspondence of phonological and grammatical words. For language-specific details of mismatches between phonological and grammatical criteria see the discussion of Fijian in Dixon (1988:21–31) and that of Yimas in Foley (1991:80–88).
In this paper, I describe a range of SVCs from Anamuxra, a Papuan language from Madang Province, Papua New Guinea, which vary in their wordhood status in ways highlighted by Dixon and Aikhenvald (2006) above. The variation in Anamuxra SVCs, which includes differences in the correlation between grammatical and phonological word, presents a challenge to current typological approaches to SVCs that generally recognise a simple binary distinction in wordhood.

The principal purpose of presenting the Anamuxra data here is to lay forth the basis for developing a more flexible and comprehensive typological account of SVC wordhood features. Another purpose of this paper is to explore the relationship between the different dimensions of wordhood and other formal and functional properties of SVCs within the language and also to examine how the Anamuxra data stands in relation to claims about the relationship between structural organisation and conceptual structure (i.e. the principle of iconic motivation).

The remainder of the paper is organised as follows: §2 gives a typological overview of Anamuxra, providing background to relevant features of the grammar of the language; §3 presents the different serial verb types found in Anamuxra; §4 considers the typological approach which classifies serial verb constructions in terms of wordhood and suggests a revision of the model as it stands to account for the Anamuxra data; §5 describes the relationship between wordhood and other formal and functional properties of SVCs and §6 summarises the main findings of the discussion.

2 Notes on Anamuxra and its grammar

Anamuxra is a member of the Josephstaal group of the Southern Adelbert Range branch of the Madang family of languages. It is spoken by approximately 1253 people (1990 census) most of whom live in hamlets and small villages to the north of Josephstaal in the north-west of Madang Province in Papua New Guinea.

Anamuxra is a predominantly verb-final language; post-verbal position is generally restricted to afterthoughts. All core and peripheral material is found before the verb. In basic transitive clauses, the subject NP and object NP can occur in either order, as shown in (2) and (3). (The positioning of core NPs is determined by principles of discourse organisation).

(2)  \[ \text{Andrew}=x \quad \text{wara-pa-ŋ} \quad n\text{-tamang-pa-ŋa.} \]
\[ \text{PN}=\text{ACC.SG} \quad \text{k.o.ant-CLFRESID-PL} \quad 3\text{s.OBJ-bite-FPST-3p.SBJ} \]
\[ \text{‘Ants bit Andrew.’} \]

(3)  \[ \text{Wara-pa-ŋ} \quad \text{Andrew}=x \quad n\text{-tamang-pa-ŋa.} \]
\[ \text{k.o.ant-CLFRESID-PL} \quad \text{PN}=\text{ACC.SG} \quad 3\text{s.OBJ-bite-FPST-3p.SBJ} \]
\[ \text{‘Ants bit Andrew.’} \]

7 The term ‘Papuan’ is applied here to refer to the group of languages found throughout Melanesia that do not belong to the Austronesian language family. Sometimes also referred to as ‘non-Austronesian’ languages, Papuan languages do not form a single genetic group. Rather they constitute a number of distinct families and isolates, put at 50–60 by Foley (1986), and fewer, perhaps 20–30, by Foley (2000) and Ross (2004).
Locative oblique NPs and adverbs tend to occur immediately before the verb as in the case of *an-i* ‘river-OBL’ in (4), while temporal NPs tend to occur sentence initially as with *aya* ‘yesterday’ in (4).

(4)  
\[ \text{Aya} \quad \text{ar} \quad \text{an-i} \quad \text{mugu-i-r.} \]  
yesterday 1d river-OBL go.down-NEAR-1d.SBJ  
‘Yesterday, we went down to the river.’

The only obligatory element of the clause in Anamuxra is the verb; and clauses containing a full complement of core NPs are less common than clauses in which core NPs are omitted such as (5), which lacks a subject NP, or (6), which lacks an object NP.

(5)  
\[ \text{Maurim-pa} \quad \text{tbu-i-na.} \]  
sweet.potato-CLFRESID bury-NEAR-2s.SBJ  
‘You bury sweet potato.’

(6)  
\[ \text{U-kura} \quad \text{kra-pa-ri.} \]  
INDEF-CLFMALE cook-FPST-3s.SBJ  
‘One man cooked (sago).’

Anamuxra is a heavily suffixing language, with some prefixing. It is primarily a head-marking language with some concomitant dependent marking. Verbs are the most morphologically complex of the word classes. The grammatical verbal word takes a single set of inflections from a set of verbal inflectional categories which include core participant marking as well as tense, aspect, mood, polarity, and switch-reference/relative tense marking. Subjects are cross-referenced on verbs in all forms except negative realis (see (7) below). Animate subjects are cross-referenced for person and number as in (2), (3) and (4). Verbs with inanimate subjects invariably take third person singular inflection. Objects are indexed on the verb by a pronominal prefix which is restricted to animate participants as in (2) and (3). Inanimate objects are not overtly marked as seen in (7).

(7)  
\[ \text{Yi-ma} \quad \text{na-wu} \quad \text{mugu-pa} \quad \text{ixr-ba.} \]  
1s.SBJ-NEG 2s-POSS house-CLFRESID see-NEG  
‘I didn’t see your house.’

Positive polarity is unmarked on the verb ((2) to (4)); negative is marked discontinuously by a suffix on the verb plus pre-verbal negative word *-ma* which takes a pronominal prefix marking person and number of the subject as in (7).

Anamuxra clauses divide into two main types: final and medial clauses. Final clauses are those that can terminate switch-reference chaining structures as in (8), which are structures that consist of more than one clause, where the identity of the subject referent in each clause is identified by morphological means as being either the same or different.

Final clauses divide into two types: final co-dependent, which correspond to counterfactual clauses (see Ingram 2001), and independent final clauses such as those in (2) to (7) and clause ‘E’ in (8) below. In addition to subject and object marking, the verbs that head final independent clauses take tense, aspect, and mood suffixes. Medial clauses represent non-final constituents of switch-reference chaining structures. They also divide into two major types: same subject and different subject. Same subject medial clauses\(^8\) are

---

\(^8\)*’Same subject’ here refers to the fact that the subject of the marked clause is the same as the subject of the following clause.*
headed by verbs that are marked just for subject person-number. Different subject medial verbs take tense and subject marking as final independent verbs plus an additional portmanteau suffix marking different subject plus relative tense: either sequential (-i) or simultaneous (-mana). The various types of medial clause are exemplified in (8): clauses B and C are same subject sequential medial clauses; while clause A represents a different subject sequential medial clause and clause D is an example of a different subject simultaneous medial clause.

(8) \[Vwa-ad-ki ay-tama-pa-ri-i\]_A\]
settlement-DEM-OBL 1p.OBJ-put-FPST-3s.SBJ-DS.SEQ \[mugu-\]_B\]
go.down-1p.SBJ
\[vwa-r makuN-tama-\]_C\]
place-OBL cargo-put-1p.SBJ \[kn-pa-\]_D\]
sleep-FPST-1p.SBJ-DS.SEQ \[kar-a-ka\]
car-NDIST-CLFRESID
\[itxa-pa-ri.\]_E\]
return-FPST-3S.SBJ

‘It put us at a settlement, and then we went down, and then put (our) cargo, and while we slept, the car returned.’

Finally, the phonological word in Anamuxra may be defined on the basis of a number of parameters including the assignment of primary word-level stress, boundary-related allophonic variation and boundary associated rules of deletion and ellipsis. While all of these are useful in determining the boundaries of the phonological word, the most consistent of these is clearly primary word stress assignment. (For details on the other parameters see Ingram 2001.) Basically, primary stress is assigned in Anamuxra to a final heavy syllable where a heavy syllable equals VV or VC; if there no final heavy syllable, stress is assigned to the penultimate syllable. Examples of words displaying final stress include:

(9) a. /aβaβ/ \[aβáβ\] ‘mother’s brother’
b. /u-taβ/ \[utáβ\] ‘INDEF-CLF
  SHELTER
’
c. /ar-i-n/ \[arín\] ‘scrape-NEAR-1s.SBJ’
d. /muN-paŋ/ \[mumbáŋ\] ‘k.o.yam-CLFRESID’

Examples of words bearing penultimate stress include:

(10) a. /ar-pa/ \[árp\] ‘tree-CLFRESID’
b. /βi-μu-ra/ \[βímúə\] ‘come-3s.SBJ-NEG’
c. /manN-aba/ \[manŋámb\] ‘banana-CLFRESID’
d. /apapura-pa/ \[apapúɾáp\] ‘butterfly-CLFRESID’

Stress associates with underlying vowels wherever possible. Thus, ‘target’ syllables (i.e. penultimate and final heavy syllables) which contain an epenthetic vowel at surface level are not immediately available to the rules of stress assignment. Rather, stress is assigned to the nearest syllable to the left containing an underlying vowel. In words where there is no underlying vowel to the left of the target syllable, the closest syllable to the right is selected.

(11) a. /βiβi-t-\-ba/ \[βíβítimba\] ‘little-CLF
  PART’
b. /id-pa/ \[índëpa\] ‘fishing spear-CLFRESID’
c. /t-Ø-na/ \[tiná\] ‘do-NEAR-2s.SBJ’
There are cases of discrepancies between the surface realisation of stress and the stress rules outlined above. One such case is found in verbs inflected for third person singular subject and NEAR tense which take antepenultimate stress. Here stress is clearly morphologically driven. Examples are:

(12) a. /dar-i-a/ [dáriya] ‘hear-NEAR-3s.SBJ’
    b. /tama-i-a/ [támiya] ‘put-NEAR-3s.SBJ’
    c. tama-pa-ri [támapari] ‘put-FPST-3s.SBJ’

This contrasts with the predictable assignment of stress in other NEAR tense forms of the same verbs with different subject person-number marking.

(13) a. /dar-i-na/ [darína] ‘hear-NEAR-2s.SBJ’
    b. /tam-i-ŋa/ [tamíŋa] ‘put-NEAR-3s.SBJ’
    c. /tur-i-n/ [tɔrın] ‘hold-NEAR-1s.SBJ’

3 Serial verb types in Anamuxra

It is well known that a single language can have more than one serial verb type (Durie 1997; Pawley and Lane 1998; Foley and Olsen 1985; Aikhenvald 1999, 2006). In Anamuxra, three SVC types can initially be distinguished based on the inflectional possibilities shown by the initial verb in the sequence: juxtaposed SVCs, medial SVCs and independent SVCs. These three are used to express a variety of meanings. They also differ in terms of their functional properties, in the marking of grammatical categories and in terms of their wordhood.

3.1 Juxtaposed SVCs

Juxtaposed SVCs comprise what is sometimes known as ‘root serialisation’ or ‘(verb) compounding’ (Aikhenvald 2006; Durie 1997). They consist maximally of two verb stems; nothing can intervene between these. The whole SVC takes a single set of inflections: object prefixing must precede the first verb in the sequence, while all suffixal material (i.e. independent and dependent morphology) occurs after the second stem. There is no suffixal material associated with the first verb in the series. For example, in (14), the verb series -tuwu-vada-, which is the head of the final independent clause, takes the third person singular object prefix n-, and a single tense marker -pa ‘far past’ and subject suffix -ri ‘3s.SBJ’.

(14) IrbaN-sx-t nurugugu-a-ki sa-t aikvara-t
    monitor.lizard-turn.into-3s.SBJ k.o.tree-NDIST-OBL be-3s.SBJ crawl-3s.SBJ
    mga-t n-tuwu-vada-pa-ri.
    come.down-3s.SBJ 3s.OBJ-hit-break^9-FPST-3s.SBJ
    ‘He turned into a monitor lizard and waited in the tree and then he came crawling down and he split her.’

^9 Tests have shown vad- ‘break’ to be a transitive only predicate when used in isolation, though, as noted by Durie ‘one cannot assume that a serialised verb will have the same argument structure in serialisation that it has when used alone’ (1997:324). The verb ti ‘hit’ in Igbo (Lord 1975), for instance, has been shown to vary in its transitivity value between its independent use and its employment within a serial verb.
Example (15) presents a negated juxtaposed SVC. Here the general negative suffix -ba attaches after the final verb stem in the series, while the negative word -ma occurs before the serial verb.

(15)  Ad-a-m n-ma sxa n-tuwu-vad-ba.
DEM-NDIST-CLF\textsubscript{MALE} 3s.SBJ-NEG again 3s.OBJ-fight-break-FPST:NEG

‘He didn’t split her again.’

In juxtaposed SVCs, the component verbs not only form a single grammatical word but also a single phonological word, with the whole construction assigned a single primary word level stress. For example, n-tuwu-vad-ba 3s.OBJ-fight-break-FPST-NEG is realised as [tuw\textsuperscript{u}β\textsuperscript{á}ndim\textsuperscript{b}a].

Juxtaposed SVCs are mostly symmetrical (i.e. both verbs come from an open class) and express several kinds of meanings. The example in (14) contains a symmetrical SVC expressing a relationship of cause and effect. In SVCs of this type both verbs are transitive; the first verb describes the type of action while the second verb focuses on the effect the action has on the object. Another example of a cause-effect SVC is given in (16).

(16)  Avn-a-xida aru-t-t xdx-a-ŋ xubaNx-a-ka
coconut-NDIST-CLF\textsubscript{TREE} large-VSF-3s.SBJ root-NDIST-PL hole-NDIST-CLF\textsubscript{RESID}
simu-ix-pa-ri.
pierce-close-FPST-3s.SBJ

‘The coconut tree grew large and the roots closed the hole.’

Example (17) presents a symmetrical juxtaposed SVC where the first verb expresses an event that results in the event specified by the second verb. In this case, the second verb is intransitive.

(17)  Nxwu maravNx-a-ki wuta-mugu-pa-ri.
3s:POSS ginger-NDIST-OBL press-go.down-FPST-3s.SBJ

‘He pressed (it) into his ginger.’

In (18) the serial verb tara-sx- ‘put.in-carry.stringbag.on.head-’ expresses sequential action. The ordering of the two verbs is iconic: the event expressed by the first verb occurs before that expressed by the second verb.

(18)  AnN-vk-pa-ng wunaN-pa tara-sx-t
water-bamboo-CLF\textsubscript{RESID}-PL plate-CLF\textsubscript{RESID} put.in-carry.bag.on.head-3s.SBJ

mg\textsuperscript{u}-t nbur-na-t wax-t ita-pa-ri.
go.down-3s.SBJ wash-EXT-3s.SBJ scoop-3s.SBJ stand.up-FPST-3s.SBJ

‘She put (the) water bamboo and plates into (a string bag) and carried it on (her) head and went down and washed (them) and scooped (water) and stood (them) up.’

Another example of a juxtaposed SVC showing sequential action is illustrated in (19). Here both component verbs are underlyingly transitive.

(19)  Arig-a-ka sis-t nxwu tutuxamk-a-ki
betelnut-NDIST-CLF\textsubscript{RESID} tear.off-3s.SBJ 3s:POSS shirt-NDIST-OBL
Andrew Ingram

\textit{uvû-tara-t mud-i-a.}
\text{twist.off-put.in-FPST:3s.SBJ come.up-NEAR-3s.SBJ}
\text{'He tore off the (hand) of betelnut and twisted (them) off and put them into his shirt and then he came up.'}

The example in (20) is a juxtaposed SVC in which the second verb indicates that the event of the first verb is carried out as a means of checking the object (see also discussion of medial SVCs).

(20) \text{Wa-ta tuwa-was-na-i aviN-vi angi-di, ...}
\text{go-2s.SBJ hit-check-NEAR:2s.SBJ-DS.SEQ good-ADV cry-3s.SBJ}
\text{'You go check by hitting (it) and if it cries well, (then) ...'}

The meanings of the juxtaposed SVCs discussed above are more or less compositional (i.e. the meaning of the SVC is determinable from the meanings of the component verbs). There are also, however, a number of forms in the language such as those in (21) that are semantically unpredictable (i.e. the meaning of the whole is not determinable from the meanings of the constituent verbs).\(^{10}\)

(21) a. \text{tagu-av-step-build.shelter-}'surround'
b. \text{tuwu-var-hit-say-}'blow'
c. \text{tama(x)-ixr-put-look-}'look up'
d. \text{v-tagu-get-step-}'walk'
e. \text{muda-tagu-come.up-step-}'climb (up or down)'

However, even in SVCs that are non-compositional there is generally a plausible link between the meaning of at least one of the verbs in the series with the overall meaning of the whole.\(^{11}\)

The non-compositionality of examples such as (21) is what distinguishes them from ‘predictable’ juxtaposed SVCs.\(^{12}\) In terms of formal properties, the two are not discernibly different.

Juxtaposed SVCs then, constitute a class of forms which consist of verb sequences comprising a single phonological and grammatical word that express a variety of meanings ranging from those that are generally predictable from the components of the series to those that are quite opaque.

\(^{10}\) Both Durie (1997) and Aikhenvald (1999, 2006) have observed that symmetrical SVCs are prone to lexicalisation.

\(^{11}\) Other Papuan languages displaying lexicalisation in verb serialisation that are discussed in the literature include Alamblak (Bruce 1984); Haruai (Comrie 1995); Kalam (Pawley and Lane 1998); Kobon (Davies 1981).

\(^{12}\) It must be acknowledged that even the ‘compositional’ SVCs are lexicalised to some extent (cf. Pawley and Lane (1998:205) on Kalam). The difference between compositional SVCs and non-compositional SVCs might be more properly thought of as a matter of degree.
3.2 Medial SVCs

In medial SVCs, the first verb takes the same affixes as the verb that heads same-subject medial clauses found in switch-reference chaining constructions (see §2, i.e. root plus subject person-number marker). The final member of the series takes the necessary final or medial marking. For instance, where the medial SVC heads a final independent clause, the last verb in the series inflects as a final independent verb. This is illustrated by the SVC in (22) where the first verb, *vyara-* ‘trail’ takes the medial form while *mudu-* ‘go up’ takes the far past tense marker -pa and the third person singular subject suffix -ri.

(22) *Xdat-pa a-ka vŋara-t mudu-pa-ri.*
post-CLFRESID NDIST-CLFRESID trail-3s.SBJ go.up-FPST-3s.SBJ
‘It (betelnut juice) trailed the pole up.’

In (23), the medial SVC heads a negated final independent clause. As with juxtaposed SVCs, the component verbs of medial SVCs cannot be independently negated: there can be just one negative suffix which attaches to the ultimate verb and one negative word which precedes the verb series.

(23) *Yam! Yiv n-ma itxa-t mugu-ba.*
mother father 3s.SBJ-NEG return-3s.SBJ go.down-NEG
‘Mother! Father didn’t go back down.’

Where the SVC heads a medial clause, the last verb takes the medial form. For example, in (24) the second verb of the SVC *mig-tr muduwa-tr* takes the third person dual same subject sequential form characterised by the addition of the suffix -tr.

(24) *Mga-tr Abringada mawa-pa taxa-tr tuwana-tr*
come.down-3d.SBJ PN sago-CLFRESID cut-3d.SBJ wash-3d.SBJ
*mig-tr muduwa-tr yiv-t-pa-ri-i Arskiwuk*
pour-3d.SBJ throw.away-3d.SBJ dark-do-FPST-3s.SBJ-DS SEQ PN
“Ab-ki xn-ba-pr” vara-pa-ri.
Q-LOC sleep-FUT-1s.SBJ say-FPST-3s.SBJ
‘They came down and cut (a) sago tree at Abringada and they pounded it and washed it and tipped it out and it became dark and Arskiwuk said “Where shall we sleep?”’

Medial SVCs, like juxtaposed SVCs, express a variety of meanings/functions. Unlike juxtaposed SVCs, however, most medial SVCs are asymmetrical: that is, one of the constituent verbs comes from a closed, or restricted set. It is this restricted verb that determines the meaning of the overall SVC. Take, for instance, the directional SVCs in (22) and (23). In these, the second verb is restricted to one of the members of the class of direction/motion verbs and imparts directional meaning, specifying the direction of the event described by the first verb.

The restricted verb is not always the second verb. For example, in inceptive SVCs such as (25), the verb *nujudama-* ‘begin’ occurs as the first verb, indicating the inception of the state of affairs expressed by the second verb.

(25) *Vanamuy-a-ki, nujudama-my simubr-i-ŋ t-Ø-ŋ.*
tulip.bark-NDIST-obl start-1p.SBJ bore-NEAR-1p.SBJ do-NEAR-1p.SBJ
‘With tulip bark, we begin boring.’
Andrew Ingram

The medial SVC in (26) is interesting in that it contains the same verbs as the juxtaposed SVC in (20). Furthermore, it has the same interpretation: the first verb indicates the means by which the object is checked.

(26) Wa-ta tuwa-ta was-na-i u-muku angi-di …
    go-2s.SBJ hit-2s.SBJ check-NEAR:2s.SBJ-DS.SEQ INDEF-CLF cry-3s.SBJ
    ‘You go check by hitting (it) and if it cries a little, (then) …’

Despite the similarities between (20) and (26), it is not the case that juxtaposed SVCs are merely reduced forms of medial SVCs; the two types are not in free variation. Most medial SVCs express relations in asymmetrical construction not found in juxtaposed constructions.

The component verbs of medial SVCs are not only distinct grammatical words; they also occur as separate phonological words, each taking discrete word level primary stress. This is in direct contrast to the unitary phonological and grammatical word status of juxtaposed SVCs. As a distinct grammatical word, the initial medial verb form takes subject person-number marking.

Clearly, the form of the non-final verb in the series stands as a contradiction to criterion ‘iii’ in (1), which stipulates that there are no morphemes characteristic of clause boundaries. However, as we have noted the identification of SVCs involves the clustering of features rather than the satisfaction of all the criteria given in (1). And, as we have seen, medial SVCs are treated as a single entity in the realisation of other categories associated with final independent verbs such as tense and negation, which are marked just on the ultimate verb of the series.

3.3 Independent SVCs

Independent SVCs are asymmetrical constructions: the initial predicate position is open, while the ultimate verb position is filled by the generic verb $t$- which, when used as a lexical verb, has a reading equivalent to English ‘do’, ‘act’ or ‘make’. All independent SVCs express habitual aspect. They are given their name because the non-final verb takes the form of an independent verb.\(^{13}\) As shown in (27), to form the positive present habitual, both verbs take NEAR tense final independent forms, while, as shown in (28), to form the past habitual, the first verb takes the NEAR tense form, while $t$- takes the far past tense form.

(27) $\text{N-xuptm-a-ka}$ $\text{ar-i-ŋ}$ $\text{t-Ø-ŋ}$.
          3s.POSS-neck-NDIST-CLFRESID scrape-NEAR-1p.SBJ do-NEAR\(^{14}\)-1p.SBJ
    ‘We usually scrape its neck.’

In the past

(28) $\text{AsaraN-tbuv-Ø-n}$ $\text{t-pa-n}$.
          tobacco-smoke-NEAR-1s.SBJ do-FPST-1s.SBJ
    ‘I used to smoke.’

\(^{13}\) While such constructions raise questions about headedness in SVCs, these issues fall outside the scope of the present paper.

\(^{14}\) Verbs vary with respect to marking of near tense. Verbs such as $t$- take a zero morpheme, others take -$i$. For details of conjugation classes in Anamuxra see Ingram (2001).
Despite the mismatch in marking present in the past habitual, the construction is interpreted as having a single overall tense value.

This is true also of negation. As illustrated in (29), in negative habitual constructions both constituents of the habitual SVC are marked with the negative suffix -ba.

(29) Yi-ma n-ba t-ba.
    1s.SBJ-NEG chew-NEG do-NEG
    ‘I don’t chew (betelnut).’

As with tense, the morphological realisation of negation illustrated here does not represent separate negation of the component verbs; the whole SVC is interpreted as having a single, unified value.\(^{15}\) This is, of course, in accordance with the defining criteria set out for SVCs in (1). That the double marking of negation here represents concordant marking only is reflected by the grammatical unacceptability of constructions such as (30a) and (30b) in which one of the constituent verbs takes the negative suffix and the other takes positive form.

(30) a. *yi-ma na-n t-ba
    1s.SBJ-NEG chew-NEAR:1s.SBJ do-NEG

b. *yi-ma n-ba t-Ø-n
    1s.SBJ-NEG chew-NEG do-NEAR-1s.SBJ

It is also reflected by the fact that otherwise there can be just one negative word per clause. Structures such as (31) are grammatical only when interpreted as biclausal constructions and not attempts at habitual serialisation.

(31) Yi-ma n-ba yi-ma t-ba.
    1s.SBJ-NEG chew-NEG 1s.SBJ-NEG do-NEG
    ‘I didn’t chew (betelnut). I didn’t.’ /*’I don’t chew (betelnut).’

Although both the main verb and t- take inflections associated with independent verbs, habitual SVCs function as a single syntactic predicate. The components form a single phonological word and cannot be interrupted by prosodic features (i.e. pauses or intonation contours) associated with distinct independent clauses. For example, \( ar\-i\-ŋ t-Ø-ŋ \) ‘scrape-NEAR-1P.SBJ do-NEAR-1P.SBJ’ is realised as [arŋtín].

This section has introduced three formally distinct SVC types, which amongst other things, differ in their status as words. (The properties of the three types are summarised in Table 3.) In the next section, the different wordhood statuses of the three Anamuxra SVC types are considered within general typological approaches.

### 4 Wordhood and a typology of serial verbs

Durie (1997) suggests an account of verb sequencing which distinguishes four types of serial verb construction on the basis of the interaction between two parameters concerned with the structural properties of SVCs, namely contiguity and incorporation. In Durie’s (1997) terms, the former deals with whether arguments, or constituents, can intervene between the constituent verbs (-contiguous) or not (+contiguous), while the latter, incorporation, is concerned with whether the constituent verbs form a single phonological

---

\(^{15}\) Compare Crowley (1987) on Paamese.
word (+incorporating) or not (-incorporating). While Durie (1997) presents ‘incorporation’ as a distinction between plus and minus phonological word, some authors have since extended the scope of this parameter to include the notion of grammatical word. Aikhenvald (1999:471), for instance, interprets the incorporating parameter as concerning ‘whether a verb sequence forms a single grammatical and phonological word or not’. Both characterisations of incorporation present a binary contrast, and as such whichever interpretation of incorporation is taken, the combination of contiguity and incorporation results in a typology comprising the four possibilities given in Table 1.

**Table 1:** Four SVC types based on contiguity and incorporation

<table>
<thead>
<tr>
<th></th>
<th>- contiguous, - incorporating</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>+ contiguous, - incorporating</td>
</tr>
<tr>
<td>III</td>
<td>+ contiguous, + incorporating</td>
</tr>
<tr>
<td>IV</td>
<td>- contiguous, + incorporating</td>
</tr>
</tbody>
</table>

According to Aikhenvald (2006), types I, II and III are widely attested cross-linguistically. Examples of I are found in Sranan (Sebba 1987); II in Jeh (Gradin 1976) and Ambae (Hyslop 2001) and III in Alamblak (Bruce 1984). The status of type IV is more controversial. Durie (1997) points to SVCs such as the example in (32) from Sakao as evidence of the non-contiguous, incorporating type. In these cases, a nominal intervenes between the two verb constituents.

(32) **Sakao** [-cont, +phon, +gramm] (Guy 1974:48)

\[
\begin{align*}
\text{ɣa-βoet-noed-p-rilam} & \quad l-a\text{day} \\
\text{3p.SBJ-pluck-coconut-PRF-come(sic) to-basket} & \\
\text{‘They plucked coconuts hither into the baskets.’}
\end{align*}
\]

There is, however, a problem with the analysis of examples such as (32) as non-contiguous. If such constructions represent instances of noun incorporation and the incorporated element in such a construction is part of the verb stem, then the two verb constituents are in fact contiguous, not non-contiguous. This position has strong implications for the four-way typology proposed by Durie. If, as appears to be the case, no examples of non-contiguous, incorporating SVCs can be found other than the incorporating types, then all non-contiguous SVCs are non-incorporating. As such, the parameter of ‘incorporating versus non-incorporating’ turns out not to be a useful distinction to make in the case of non-contiguous SVCs. Consequently, rather than the symmetrical four way typology Durie proposes, we are left with three distinct types: non-contiguous versus contiguous incorporating versus contiguous non-incorporating.

---

16 Note that within the general interpretation of contiguity that is followed by Durie (1997), Aikhenvald (1999) and others it is the presence of the nominal in (32) that makes the construction non-contiguous, not the presence of the affixes *p* or *ri*. It is argued inflectional or derivational suffixes are not considered to represent a break in contiguity. Obviously, if such affixes did represent a break in contiguity, then we would have many more cases of non-contiguous serialisation on our hands. It would also mean that the parameters of phonological word and grammatical word would be relevant in the context of non-contiguous serialisation.
There is another problem with the typology represented in Table 1 that is highlighted by the Anamuxra data. We saw in the previous section that it is possible to distinguish three types of SVC: medial SVCs, juxtaposed SVCs and independent SVCs. Unfortunately, Durie’s model distinguishes just two types of serial verb within the Anamuxra set: type II (medial SVCs) and type III (which encompasses both juxtaposed SVCs and independent SVCs).

The mismatch between the structural possibilities distinguished in the typology shown in Table 1 and those evidenced in Anamuxra SVCs, where juxtaposed SVCs and independent SVCs are under-differentiated within the typology, is due to the fact that Durie’s model allows for just one type of word, while Anamuxra SVCs display different wordhood properties. Specifically, as was shown in §3, juxtaposed SVCs comprise both a single phonological word and a grammatical word, whereas the constituent verbs of an independent SVC form a single phonological word but distinct grammatical words.

That a language has an imperfect correspondence between phonological word and grammatical word is not unusual from a cross-linguistic perspective, as has most recently been illustrated in Dixon and Aikhenvald (2006). However, while the possibility for such a mismatch to occur in serial verb constructions has been flagged by Aikhenvald (2006), it has not been incorporated into general frameworks dealing with the classification of structural types of serial verbs.

In order to account for the Anamuxra data within Durie’s general framework, then, we need to have ‘grammatical word’ and ‘phonological word’ as distinct parameters. Recall that we suggested earlier that wordhood, or ‘incorporation’ in Durie (1997) and Aikhenvald’s (1999) terms, was only applicable in the context of contiguous serialisation. This was because in all cases of non-contiguous SVCs attested cross-linguistically the constituent verbs do not constitute a single grammatical and/or phonological word. Thus, our revised typology of SVC types based on contiguity, phonological wordhood and grammatical wordhood presents five possibilities, which are given in Table 2.

<table>
<thead>
<tr>
<th>Table 2: Revised typology of SVC types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia</td>
</tr>
<tr>
<td>Ib</td>
</tr>
<tr>
<td>Ic</td>
</tr>
<tr>
<td>Id</td>
</tr>
<tr>
<td>II</td>
</tr>
</tbody>
</table>

Of these, Anamuxra SVCs represent types Ia (independent SVCs), Ib (juxtaposed SVCs) and Id (medial SVCs). Obvious points of typological interest are whether all the types are attested cross-linguistically and what their distribution is like. While individual descriptions of serial verbs vary greatly in their reporting of matters of wordhood, which makes it difficult to obtain an extensive or accurate picture of the distribution of the various types listed in Table 2, a preliminary survey of the literature suggests that examples of all types can be found. For instance, Ic is found in Paamese as in (33); II is found in Ambae as in (34).
Another question that needs to be asked of the revised typology is what significance the different SVC types proposed hold as discrete construction types. However, the answer to this question requires a detailed investigation that is beyond the scope of the current discussion. (It is also a question difficult to answer given the lack of uniform detail in the description of wordhood properties of SVCs across individual languages.) I use the remainder of this paper to consider how the dimensions of wordhood correlate with other formal and functional properties in Anamuxra and how they reflect upon the notion of iconic motivation.

5 Wordhood and other formal and functional properties

The combination of the discrete dimensions of phonological and grammatical words allows us to maximally distinguish between the three types of serial verb in Anamuxra using the cross-linguistically useful parameter of wordhood. In this section, we reflect on how wordhood relates to other formal and functional properties of serialisation in Anamuxra. I begin by presenting a summary of the major formal and functional properties of the three SVC types.

Table 3: Properties of three types of serial verb

<table>
<thead>
<tr>
<th>Properties</th>
<th>Juxtaposed</th>
<th>Independent</th>
<th>Medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vs=separate phonological words</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Vs=separate grammatical words</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>symmetrical/asymmetrical</td>
<td>both</td>
<td>asymmetrical</td>
<td>asymmetrical</td>
</tr>
<tr>
<td>verb from closed class</td>
<td>V2</td>
<td>V2</td>
<td>V1, V2</td>
</tr>
<tr>
<td>lexicalised</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>grammaticalised</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>functions</td>
<td>cause-effect</td>
<td>habitual</td>
<td>directional-inceptive checking manner</td>
</tr>
<tr>
<td>marking of grammatical categories tense</td>
<td>after V2</td>
<td>both (single value)</td>
<td>after V2</td>
</tr>
<tr>
<td>subject</td>
<td>after V2</td>
<td>both (agree)</td>
<td>both</td>
</tr>
<tr>
<td>negation</td>
<td>after V2</td>
<td>both (single value)</td>
<td>after V2</td>
</tr>
</tbody>
</table>
One observation we can make about serial verbs in Anamuxra in terms of the dimensions of wordhood concerns the respective formal and functional markedness of the three types, particularly that of independent SVCs. In both juxtaposed SVCs and medial SVCs the phonological and grammatical criteria align; in the former, to represent a single phonological and morphological word; in the latter, to represent two separate phonological and grammatical words. In contrast, independent SVCs do not display a symmetrical correspondence between phonological and grammatical word. As the only SVC to do so, they are in a sense, unusual, or marked. Their unusual, or marked status, is reflected in other formal and functional characteristics. Formally, independent SVCs show the most divergent properties: they are the only construction to have marking of tense on both verbs, with the possibility of a mismatch in form in the past habitual form tense marking; they are also the only ones to have negation marked on both constituents. Functionally, independent SVCs are also the most restricted of the three types, being used to express just one function, habitual aspect, in contrast to the range expressed by the other two SVC types.

The dimension of grammatical wordhood also allows for certain generalisations about the behaviour of the different SVC types in Anamuxra. For instance, it can be observed from Table 3 that SVCs consisting of distinct grammatical words (that may or may not also be separate phonological words, that is, medial SVCs and independent SVCs) do not lexicalise in Anamuxra, while SVCs that are both a single grammatical word and a single phonological word (namely juxtaposed SVCs) do show a tendency towards lexicalisation. By contrast, the parameter of phonological word appears to hold little significance in accounting for other formal or functional characteristics of SVCs in Anamuxra. SVCs which comprise a single phonological word (i.e. juxtaposed SVCs and independent SVCs) share little with each other in other formal or functional terms that one or both do not also share with SVCs that do not form a phonological word (i.e. medial SVCs). This is illustrated in several ways. As we have just seen, juxtaposed SVCs tend towards lexicalisation, whereas independent SVCs like medial SVCs show a greater tendency toward grammaticalisation. Juxtaposed SVCs are more likely to be symmetrical, whereas independent SVCs are like medial SVCs in being asymmetrical. Conversely, juxtaposed SVCs overlap slightly with medial SVCs in the functional relations they express: sharing as they do the ‘checking’ meaning.

Finally, Anamuxra SVCs offer further confirmation of cross-linguistically observed correlations between form and function/conceptual structure which have been explained in terms of the principle of ‘iconic motivation’. In short, as part of the principle of ‘iconic motivation’, boundedness in surface structure is seen to correlate with tightness in conceptual structure, where closeness in surface structure tends to be motivated by conceptual closeness (cf. Haiman 1985; Kirsner 1985; Aikhenvald 1999). With respect to SVCs, it has been found that in languages with more than one type of SVC, the closer the verb roots are in surface structure, the more they tend toward grammaticalisation or lexicalisation (Aikhenvald 1999:489).

As illustrated in Figure 1, the three Anamuxra SVC types form a continuum from most tightly bound to least tightly bound with juxtaposed SVCs the most tightly bound, comprising single phonological and single grammatical words; and medial verbs the least tightly bound: their constituent verbs forming separate phonological and grammatical words. Independent SVCs are between these two, being tightly bound phonologically, but less so grammatically; containing as they do, distinct grammatical words.
In terms of SVCs in Anamuxra we find that the most tightly bound SVC types do tend towards lexicalisation. Interestingly, however, the SVCs that show the greatest tendencies towards grammaticalisation are not the most tightly bound of the three SVC types.

6 Concluding remarks

This paper has presented three types of SVCs from Anamuxra, distinguished in terms of wordhood: juxtaposed SVCs (single phonological and grammatical word); medial SVC (separate phonological and grammatical word); and, independent SVC (single phonological word but separate grammatical words). That independent SVCs require the separation of grammatical and phonological criteria for defining wordhood is significant in the context of current typological approaches to SVC wordhood which to date have offered a simple binary distinction between single word versus multiple word. Notably, the distinction between phonological and grammatical criteria required for the specific case of independent SVCs in Anamuxra is on a more general level consistent with approaches to wordhood such as that articulated most recently by Dixon and Aikhenvald (2006) which have developed in response to the recognition of variation in wordhood criteria within numerous languages. Obviously, further cross-linguistic work on serial verb constructions that pays close attention to all aspects of matters of wordhood, especially prosodic properties, is required to provide us with a clearer picture of the extent to which the distinction of grammatical and phonological criteria for wordhood serves as a typologically valid contrast in cross-linguistic studies of serial verb constructions.

References


—— 2002, The functions of the Anamuxra generic verb -t- and its reflexes. MS. RCLT, La Trobe University.


Complex predicates and complex sentence types in Lavukaleve

ANGELA TERRILL

Some of the work for which Andy Pawley is best known has been concerned with clause types and clause linkage, with an emphasis on functional motivations for different types of clause linkage, and on culturally-dependent expression of ideas using different types of clauses. This work has ranged from universal to language-specific requirements for the well-formedness of various clause types. Also important in much of Pawley’s work is the idea that to describe a language properly one needs to not just describe what the available structures in the language are, but how people know when to use which structure; it is this latter knowledge that constitutes fluency in a given language. Pawley has been concerned in particular with the role of idiomaticity, and with the correct location of the division between the grammatical and the lexical (Pawley and Syder 1983; Pawley 1987, 1993; Pawley and Syder 2000). One of Pawley’s students, Cynthia Farr, has also used this type of approach to language description (Farr 1999). From these ways of approaching description of linguistic structure, and the description of clausal structures in particular, my own thinking about Lavukaleve clause types has been shaped. In Lavukaleve there are a number of different complex predicate and complex sentence types; if one assumes a functional niche for each clause type, then it is not enough to simply describe the morphosyntax of each clause; one also has to describe the function of each clause type. In this paper I try to examine some of the functional motivations behind Lavukaleve speakers’ choices of which complex predicate or sentence type to use, in a small corpus consisting largely of narrative monologues.¹

1 Background

Lavukaleve has one main, but not sole, way of joining two or more verbs into complex predicates, namely serial verb constructions (SVCs) (verb compounds form another, restricted, way of forming a complex predicate with two verbs; and complex predicates can also be formed with the habitual auxiliary, and with verb-adjunct constructions, which both involve a non-verbal element in construction with a verb). Further, there are three most common ways of making complex constructions to join either simple or complex predicates:

¹ This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). Thanks to three anonymous reviewers and the editors of this volume for valuable comments on an earlier draft of this paper.
these are coordinate clauses, cosubordinate clauses (i.e. clause chains utilising medial and final verbs), and subordinate constructions. In Terrill (2003, 2004) I described the structural (i.e. morphosyntactic) make-up of each of these complex predicate and sentence types, but these descriptions were almost exclusively concerned with the morphosyntactic description of each of the clause types, and only commented briefly on the functional differences between these structures. In the first descriptive grammar of an otherwise unknown language, this was perhaps a defensible approach. But now with a few more years’ hindsight it is interesting to attempt a slightly deeper understanding of these structures. While the morphosyntactic description remains unchanged, this paper aims to provide preliminary notes towards understanding the discourse functions of each complex predicate and complex sentence type, and under what circumstances a speaker would be likely to choose one over another.

Data for this paper come from a corpus of 62 recorded narrative monologues, as well as guided elicitations, translated church service materials, and conversation, collected between 1996 and 2003 by the author.

1.1 Structure of complex sentences in Lavukaleve, with a note on complex predicates

As discussed in Terrill (2003:Ch.15, 16), Lavukaleve has independent and dependent clauses. Of the dependent clauses, there are two main types: subordinate clauses (adverbial, purposive and relative); and cosubordinate clauses (i.e. clause chains with medial verbs). These categories are all defined on the basis of morphosyntactic properties; each of these clause types is expressed by means of suffixation, and, in some cases, by changes in the type of participant marking system used (Terrill 2003, 2004).

As the following examples show, they can be used in what seem to be very similar contexts. SVCs are shown in bold and cosubordinate and subordinate clauses are marked with square brackets.

Serial verb construction:

(1) *Ami mina taotao o-ma vo-ham?*  
who(M) thing(F) horrible 3sf.OBJ-take come-PURP  
‘Who would want to bring a horrible thing like that?’

Subordinate clause:

(2) *[Kemus ga e-lo-fou-ge] kini o-foa.*  
rope(N) SG.N.DEF 3sn.OBJ-3d. SBJ-put.on-ANT ACTN 3s.SBJ-go.down  
‘The two boys put down the rope.’ [lit. The two having put the rope, it went down]

Cosubordinate clauses (clause chain):

(3) *[A-ma-re] vela ma-me.*  
3sm.OBJ-take-NF go 3p.SBJ-HAB  
‘They would take it [to him].’ [lit. Taking it, they would go]

Coordinate construction:

(4) *[Maki na a-o-igu], aka /*  
nut.pudding(M) SG.M.DEF 3sm.OBJ-3s.SBJ-take.out then /
Complex predicates and complex sentence types in Lavukaleve

[suma-vil enga va koi vo-ma-re],
taro-PL three PL.DEF also 3p.OBJ-take-NFIN
‘She put the nut pudding in, then also took three taros …’

The examples above are selected to highlight the fact that all of these structures can have similar functions; in particular the function of joining predicates in chronological order in order to portray a sequence of events. Anterior adverbial clauses typically express events in a chronological sequence, as do clause chains. So do coordinated clauses (i.e. sequences of independent clauses). In addition, SVCs also express events in logical or chronological sequence. Further, the epi-syntactic phenomenon of tail-head linkage serves to unite two morphosyntactically separate yet chronologically ordered events. But why does Lavukaleve have so many ways of talking about events occurring in a chronological sequence? Different means of expression suggest different functions distributed among these clause types. This paper will take each clause type in turn and try to understand something of its function, in order to be able to say more than just that the clauses express chronologically linked events.

With the relatively small numbers of tokens of each clause type it is probably not useful to cite the exact figures, but relative frequency of each clause type may be worth looking at. Therefore, in a subcorpus of three narrative monologues (753 sentences in all), the relative frequency of each complex sentence type was noted. More than one construction can occur within a single sentence; for example a medial clause in a clause chain could itself be based on a SVC. Also, series of cosubordinate, subordinate and coordinate clauses occur within the same sentence. Cosubordinate constructions were found to occur the most frequently, followed by subordinate clauses, coordination, then SVCs.

1.2 Examination of the role of serial verb constructions (SVCs)

SVCs in Lavukaleve consist of two, or rarely three or more, verbs in juxtaposition. All of their morphology is held in common, including their subjects and, if transitive, their objects, and all adjuncts. Markers of these categories are expressed only once in the construction, apart from objects, which are marked on each transitive verb of the construction. All morphology and adjuncts are shared across all the verbs of a SVC (Terrill 2003:373). As well as only expressing morphological categories once, SVCs differ further from complex clause joining structures in that the verbs of a SVC always occur in juxtaposition, without overt marking to signal their relationships. Other complex structures require overt marking to express the relationships of verbs and predicates. Verbs in a SVC act like a single predicate in that they have most morphological marking only once across the predicate. They act like separate predicates in their marking of objects, which occurs once on each transitive verb in the predicate, and in the fact that adjuncts can intervene between the verbs.

In (5), the two verbs share a subject, which is marked once, on the second verb:

(5)  
Fele-re  kini  feu  o-fufu.
  return-NFIN  ACTN  go.inland  3s.SBJ-lie.down
‘She went back inland and lay down.’
  [lit. Returning, she went inland and lay down]

In the following example, the first verb is transitive, and bears an object prefix, while the second, intransitive verb, bears an imperative suffix which has scope over both verbs.
(6)  

\[ E-\text{ma} \quad \text{vula-ma!} \]

3sn.OBJ-take  come-DUR.IMP.SG

‘Bring it!’

In (7), the SVC consists of four verbs, all transitive and all bearing object prefixes referring to the shared object argument (while object marking is obligatory, subject marking is not).

(7)  

\[ \text{Tamal!} \quad \text{Mel} \quad \text{kini} \quad \text{havu-v} \quad \text{vo-ha} \quad \text{vo-ta} \]

men.DU 1di  ACTN  ngali.nut-PL 3p.OBJ-pick 3p.OBJ-pound

\[ \text{vo-fo'foiri} \quad \text{vo-fa} \quad \text{vo-na}. \]

3p.OBJ-pummel 3p.OBJ-bake 3p.OBJ-in

‘Hey guys! Let’s go and pick some ngali nuts and pummel them and pound them and bake them.’

The first verb of most SVCs is a motion verb; in many cases the second verb is also a motion verb, thus in many cases the SVC consists of two motion verbs. Examples include the following:

- **hau veo** ‘go.shorewards arrive’ “reach the shore”
- **vau ao** ‘go.seawards go.in’ “go in seawards”
- **fele vo** ‘return come’ “come back”

Of the minority of SVCs which do not have a motion verb as their first verb, examples include:

- **ma vau** ‘take go.seawards’ “take seawards”
- **lai i** ‘tell do’ “tell”

Motion verbs also appear frequently as second verb. Second verbs are otherwise made up largely of stance verbs and verbs of induced location (e.g. put, push, throw, etc.). Examples include:

- **ae fifi** ‘go.up sit’ “sit up”
- **ae fale** ‘go.up stand’ “stand up”
- **sou fifi** ‘rise sit’ “sit up”
- **vau ko** ‘go.seawards throw’ “throw seawards”

Many of the SVC tokens consist of a stable pair of verbs, which form common collocations. ‘Take–come’ is one such collocation, as is ‘go seawards–go in’ (to describe walking seawards to the beach, then getting in a canoe). Such common collocations form the majority of SVCs. The fact that many of the combinations are repeated suggests that these SVCs are to some extent idiomatised. Such expressions represent both common daily occurrences and the correct way to describe them. SVCs represent a way to link multiple events into a single, idiomatic chunk. The fact that many of these SVCs are idioms makes one wonder if they get structurally codified as well; and in fact they do. This is the most likely origin of the verb compound. Verb compounds were described in Terrill (2003) as a special type of very tightly constrained verb sequences, which work together to form a verb compound. They differ from SVCs in that morphologically they act exactly as a single verb, whereas SVCs act in some ways as a single verb (i.e. in their morphological marking, marking morphological categories just once per SVC) and in other ways as a sequence of verbs (e.g. in their separate marking of the object for every transitive verb of
the SVC). In addition, some elements can occur between verbs of a SVC, whereas nothing can occur between verbs of a verb compound. There are only three verb compounds in the language; *ke vauri* ‘push – go seawards’; *ke feuri* ‘push – put inland’; and *sou faleri* ‘rise – stand’. These verb compounds give the impression of having arisen from SVCs, gradually becoming frozen expressions through being common collocations. It is tempting to see a cline from SVCs to verb compounds, but actually the morphological expression of each does not differ in a cline-like manner, but is rather sharply distinguished. It is probably safe to assume though that there is a historical relationship between SVCs and verb compounds, and, further, that it is the most common SVC combinations which get frozen into verb compounds. It is not by chance that two of the three verb compounds contain one of the extremely ubiquitous motion/direction verbs *vau* ‘go seawards’ and *feu* ‘go inland’. Further motion/direction verbs include *hau* ‘go shorewards’, *ae* ‘go up’, *vo* ‘come’ and *ve* ‘go’, and these verbs form a large proportion of verbs in SVCs in Lavukaleve. It is possible that some of these motion verbs forming common SVC combinations in Lavukaleve today might follow the same path towards verb compounds in the future.

2 Clause joining constructions

As stated above, there are in Lavukaleve independent, coordinated clause joining constructions, and two types of dependent clause joining constructions: subordinate (of relevance here are the subordinate adverbial clauses) and cosubordinate, or clause chaining constructions. Their relationship is outlined in Figure 1.

Lavukaleve clause types

```
  independent
    ↓
  dependent
    ↓
  subordinate
    ↓
  adverbial
  ↓
purposive
  ↓
relative
```

Figure 1: Clause types in Lavukaleve (Terrill 2004).

In earlier work I noted that coordination was ‘not the preferred way of joining clauses’ in Lavukaleve, and that much more common was joining clauses using subordination (Terrill 2004:440). I would like now to look closely into this claim, looking in particular at the circumstances under which each clause type typically gets used. Each of the relevant clause types will be discussed in turn.

2.1 The role of coordination

Clausal coordination is not as common as other types of clause linkage, including subordination and cosubordination. However, a closer look at coordination shows that
differences emerge with different speakers: some speakers use coordinated clauses more frequently than others. There could be a number of possible reasons for this, apart from individual stylistic differences, including different genres, register, age, etc. A small corpus does not afford a reliable way to estimate the power of these and other possible factors. However, it can be said that in general coordination is rather less frequent than cosubordination and subordination.

If a speaker does use coordination, a coordinator must be employed. The most common coordinate structure is achieved by means of the coordinator *aka* ‘then’ (in bold in the following examples). This can coordinate clauses or indeed whole sentences, as in (8) and (9).

(8) \[Aka-ri\ sia-nun ta fi\] *aka*, \[oiva / mina /\nthen-PR do-DUR just 3sn.FOC then other.NTRL.PL / thing(F) /\nna’nug roa a-ma-lugu].
thought(M) one.SG.M 3sm.OBJ-3p.SBJ-think
‘It went on just like that, then the others – um – had a thought.’

(9) \[Nerea-la-v gala kelea-nun kelea-nun ta\] *aka* /\nbe.close-EXT-PL there walk-DUR walk-DUR just then /\n[kafol roa a-o-le].
mangrove.crab(M) one.SG.M 3sm.OBJ-3s.SBJ-see
‘He went close to shore, then he saw a mangrove crab’

*Aka* also plays a role in a much more complex situation involving multiple juncture-marking. This is described in §2.4.

It is possible that speakers tend to prefer other ways of linking clauses rather than via coordination because coordination does not provide an explicit semantic link between the joined clauses, in the way that for example subordinate adverbial or cosubordinate clauses do.

### 2.2 The role of cosubordinate clauses (clause chains)

Clause chains are built from one or more medial clauses, followed by a final clause. The verbs of medial clauses mark only their objects and their cosubordinate status (with one of three suffixes, -\textit{re} non-finite, -\textit{vel} completive, and -\textit{vele} successive). They cannot mark their subjects or other TAM categories apart from the causative suffix.

When describing Lavukaleve’s cosubordinate clauses (i.e. clause chains) it is tempting to choose example sentences which illustrate what, reading classical examples of these constructions from other Papuan languages, one could take to be the genius of the construction: that is, a long chain going on for many clauses, ideally describing some supremely culturally appropriate event series, with each verb breaking down the event into ever more culturally-loaded scenes. Unfortunately, while languages like Kalam (Pawley 1987, 1993) lend themselves readily to this type of example, such examples are few and far between in Lavukaleve. For the sake of completeness, here are a few of the best examples of this kind in my Lavukaleve corpus (cosubordinate (medial) verbs here are all marked with -\textit{re} ‘non-finite’). Medial clauses are bracketed and final verbs are in bold type.
Complex predicates and complex sentence types in Lavukaleve

(10) *Hano ika ngoa-nun / [oinala karu na* then there stay-DUR / other.NTRL.M.DU* possum(M) SG.M.DEF

*a-kiu-re, / [a-lo-ri-re],
3sm.OBJ-kill-NFIN 3sm.OBJ-finish-CAUS-NFIN

*[a-se-re], [a-fa-re], ika ngoa-nun.
3sm.OBJ-butcher-NFIN 3sm.OBJ-bake-NFIN there stay-DUR
‘Then, the other two stayed on there, killed the possum, butchered it, baked it, and stayed there.’

(11) *[O-ma-re], [ma-taila vo-re] / [ui e-u-re] [lo-re], ma-losi-v
3sf.OBJ-take-NFIN 3p.POSS-house-LOC* come-NFIN / food(N)

*vo-na mina* o-ma-ho.
3p.OBJ-in thing(F) 3sf.OBJ-3p.SBJ-put.inside
‘Taking it to their houses, and having finished eating, everyone put something in their food baskets.’

If this is the ideal of clause chain constructions, actually it is reached rather seldom in Lavukaleve. In fact, most of the examples of clause chains in Lavukaleve consist of only one medial verb in the chain. The exact breakdown by number of medial verbs in a subset of the data consisting of three narrative monologues is as follows:

<table>
<thead>
<tr>
<th>Medial clauses per chain</th>
<th>% of whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>117</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

(There were no chains of more than five medial verbs in this subset of the corpus.)

Therefore, the most characteristic clause chain for Lavukaleve actually looks something like this:

(12) *[A-le-re], a-o-ma.
3sm.OBJ-see-NFIN 3sm.OBJ-3s.SBJ-take
‘Seeing it, he took it.’

In clause chains there is always a temporal element. There is always an assumption of events linked chronologically in order of occurrence. The examples above show this clearly. These long strings are always used to express culturally appropriate sequences of events; that is events which follow on from each other if one is performing normal natural tasks of everyday life. For example in (10), there is a natural sequence of events by which possums are hunted, butchered and eaten, and this type of natural chain of events lends itself to a chain of clauses; hence, the longer clause sequences in this type of event description.
But the counts of number of clauses per chain give the lie to the idea of Lavukaleve having the classical Kalam-like endless event chains. Clearly, chains are normally used to describe short events too. But what exactly are clause chains doing, if not stringing together long sequences of events? In fact, short clause chains in Lavukaleve very often serve to separate out aspectual information from content-information. Typically, content information comes in the medial clause, with aspectual information in the final clause (medial clauses are bracketed, final verbs in bold). For example:

(13) \([\text{Mina ro nua-re}]\) fi \(\text{o-lei}.\)
    \(\text{thing(F) one.SG.F be.amiss-NFIN 3sn.FOC 3s.SBJ-exist}\)
    ‘Something was wrong.’ [lit. Something is amiss it is]

(14) \([\text{Vo-vea-re}]\) fi \(\text{ngo-lei ke}.\)
    \(3p.OBJ-know-NFIN 3sn.FOC 2s-exist EMPH\)
    ‘You know how they are.’ [lit. Knowing them you are]

(15) “\([\text{E-liki-re}]\) fi \(\text{le-lei} \ a-e-re-ge.\)
    \(3sn.OBJ-want-NFIN 3sn.FOC 1de-exist 3sm.OBJ-SBD-say-ANT\)
    ‘“We two want it” he said.’ [lit. Wanting it, we are]

(16) \([\text{Maki na a-ta-re}], \ [\text{seluka-vil nut.pudding(M) SG.M.DEF 3sm.OBJ-pound-NFIN sore-PL}]
    ma-gogomet va vo-ma-re], \ [nago la 3p.POSS-scab(PL) PL.DEF 3p.OBJ-take-NFIN mortar(F) SG.F.DEF]
    o-koli-n vo-ho-re], aka-ri sia-ne.
    \(3s.POSS-interior-LOC 3p.OBJ-put.inside-NFIN then-PR do-IPF\)
    ‘Pounding the nut pudding, taking the scabs off her sores, putting them inside the mortar, like that [she] did.’

(17) \([\text{Vala sia-re} \ [\text{ali inu / visave-a-re}] \ [\text{kiu-re}] \ o-lei}
    how do-NFIN man(M) 2s / mourn-INTR-NFIN die-NFIN 2s-exist
    ‘Why are you so terribly upset!’ [lit. Why mourning dying you are!]

In the first three examples above, content information is followed by aspectual information. Content information can be expressed in the medial clause, with aspectual information encoded by \(\text{lei ‘exist’}\) in the final clause, as in (13), (14) and (15). A similar situation occurs in example (16), with content information expressed in the medial clauses and aspectual information in a generic verb ‘do’ in the final clause. In (17) the first two medial clauses have content information encoded in ‘mourn’ and ‘die’, and the third, final verb has aspectual information encoded by \(\text{lei ‘exist’}\).

In keeping with the role of medial verbs as content holders and final verbs as aspectual or semantically more general, the following example shows, in the second clause chain, a medial verb \(\text{fele ‘return’}\) expressing path and motion, and a semantically more empty final verb \(\text{ve ‘go’}\) expressing generic motion.

(18) \(\text{O-ko-re fo’sal va vo-vala-re lo-re,}
    3sf.OBJ-throw-NFIN ffish(PL) PL.DEF 3p.OBJ-pull-NFIN finish-NFIN
    aunio-n nun aka-ri / [\text{fele-re}] ma-ve.
    evening-LOC from then-PR / return-NFIN 3p.SBJ-go
    ‘They threw in the lines, and having pulled some fish, evening came and returning they went.’
This tendency of content and aspect or generic motion or action to be split across clauses does not only occur within clause chains. Examples (19) and (20) are SVCs in which the verb of the first clause, akuku in (19), contains the content information, with the second verb doing little more than hosting subject and the future tense marker. Splitting the parts of the full intended meaning up allows the speaker to express focus on the first, contentful, section, using a focus marker with scope over the first verb.

(19) A-kuku fi a-i-re.
   3sm.OBJ-carry.on.back 3sn.FOC 1s.SBJ-do-FUT
   ‘I’ll carry it.’ [lit. Carry it I will do]

The following construction serves the same purpose too: allowing focus marking over the content verb, and using the generic, final verb as host for subject and tense marking.

(20) Inu koi ngo-u fi a-i-re.
   2s also 2s-eat 3sn.FOC 1s.SBJ-do-FUT
   ‘I’m going to eat you too.’ [lit. Eat you too I will do]

So this represents another and very frequent way of joining clauses. Structures like these verge on the idiomatic: they form a conventionalised way of breaking down sequences of events. In summary, cosubordinate (clause chain) structures are often used as sequencing devices, which are used to space out information, in particular to separate aspectual information from the more strictly contentful verbal information.

### 2.3 The role of subordinate constructions

There are three types of subordinate clauses: adverbial, relative and purposive. Here we will discuss only the adverbial, as they fall within the scope of our starting observation of the multiple ways that Lavukaleve has of joining clauses in chronological order in order to portray a sequence of events. Subordinate adverbial clauses always precede the matrix clause and are of three types: anterior, marked by suffix -ge, potential, with suffix -le, and surprise, with suffix -meon. As well as these suffixes, subordinate adverbial clauses are marked by a prefix e- and a split-ergative prefixing system, with third person subjects following an ergative-absolutive marking pattern, and first and second person subjects following a nominative-accusative pattern. Of the three types, the anterior is by far the most common, with the surprise suffix occurring in the corpus only in a handful of cases. Anterior subordinate clauses can be translated as something like ‘Having done X, then Y happened’. This is an extremely common way of presenting events in chronological sequence. Subordinate adverbial constructions do not typically have more than one subordinate clause in sequence. Unlike clause chains, subordinate adverbial structures are not a way of spacing out information and dividing the load between aspectual and content information. They do provide a way of joining one event to another, and the events are always stated in chronological order, but the subordinate adverbial clause provides a backdrop, or scene setting, for the matrix clause in which it is embedded. Common scenes in which subordinate constructions are used include (21), in which the first clause (‘darkness having fallen’) provides the backdrop for the main clause (‘she emerged and stood up and went seawards’).

(21) [Koro ga e-fau-ge] aka oia /
darkness(N) SG.N.DEF 3snO.SBD-happen-ANT then other.NTRL.SG.F /
In (22), again the anterior clause ‘the day having come’ provides background to the main event of the main clause ‘went seawards’.

(22) ... le roge o-vea koi / [e-vea-ge],
day(N) one.SG.N 3s.SBJ-emerge also / 3snO.SBD-emerge-ANT
koi lo-vau.
also 3d.SBJ-go.seawards
‘... one day came, then it having come, the two also went seawards.’

In (23), events are again stated in chronological order.

(23) [A-e-huru-ge] koi, a-ham / maki
3sm.OBJ-SBD-go.inside-ANT also 3sm.OBJ-for / nut.pudding(M)
na a-ta fi lo-me. Koi.
SG.M.DEF 3sm.OBJ-pound 3sn.FOC 3s.SBJ-HAB also
‘Him having gone inside, she makes nut pudding for him. Again.’

Subordinate clauses can also, less commonly, be involved in chains themselves, with subordinate rather than cosubordinate medial verbs: that is, a series of subordinate verbs can occur, followed by a final main verb.

(24) (The other man stood up, but he took the wrong path. He walked and walked and walked, and then his little fish got smelly [went rotten].)

[O-e-kivi-ge], [le ga
3sf.OBJ-SBD-be.smelly-ANT day(N) SG.N.DEF
e-vea-ge], o-loolo kini
3sn.OBJ.SBD-emerge-ANT 3s.POSS-straight ACTN
o-taila-n lavea-m.
3s.POSS-house-LOC appear-SG.M
‘Them having gone rotten, then next day having come, he went straight home.’
[lit. Straight away he appeared in his house]

2.4 Higher-level structures

Described above are some of the basic structures which speakers have recourse to in packaging information to express sequences of events in various ways. However if one moves away from the straightforward examples given above, the picture rapidly becomes much more complex.

For example, subordinate structures and clause chains both have overt morphology marking the nature of the junction between the marked clause and the matrix clause, but they can also occur with a coordinator between the two clauses. This can be thought of as multiple juncture marking. An example with coordinator aka:

(25) Veo-re, fi aka e’rau-re o-feu.
arrive-NFIN 3sn.FOC then jump/fall-NFIN 3s.SBJ-go.inland
‘Arriving, jumping out, he went inland.’
One speaker in particular, Clement Oiva, uses a particular structure extremely frequently, consisting of a subordinate-marked clause followed by aka, which other speakers use rather infrequently. For example, in (26) there is a perfectly coherent subordinate clause ofafage, which however also has aka ‘then’, a coordinator, appearing where it should appear between the two clauses, as if it were functioning as coordinator:

(26)  
\[
\text{[O-e-fau-ge], aka hide o-re “Tam!”} \\
3sf.OBJ-SBD-happen-ANT then thus 3s.SBJ-say man
\]
‘It happening, then [one man] said “Man!”’

Also, coordinator aka sometimes occurs between medial verbs, adding extra coordination leverage in a chain. In this type of situation aka seems to be functioning more as a discourse structuring element rather than just as a coordinator:

(27)  
\[
\text{Aka iutia-re-ne [ngoa la-e-me-ge]} \\
\text{then look.on-NFIN-IPF stay 3sm.OBJ-SBD-HAB-ANT}
\]
\[
\text{[oinala kafol na a-u-re]} \text{[lo-re]}, \\
other.NTRL.mdu mangrove.crab(M) SG.M.DEF 3sm.OBJ-eat-NFIN finish-NFIN
\]
\[
\text{aka [o-rioko la o-ma-re], lai-n} \\
\text{then 3s.POSS-bag.of.crab(F) SG.F.DEF 3sf.OBJ-take-NFIN top-LOC}
\]
\[
\text{fi o-kaego.} \\
3sn.FOC 3sf.OBJ-put.on.top
\]

‘[Him] looking on, as the other two ate the mangrove crab, he put the crab up on top’ [lit. Then [him] keeping looking on, the other two eating up the mangrove crab, then taking its bag, put it up on top.]

Multiple juncture-marking of this sort is rather a challenge to the idea of a static morphosyntactically defined clause typology. It seems that rather than choosing one clause juncture type, speakers can also use rather looser structures which could be interpreted in a number of different ways. Note that speakers do not stack morphology; they do not put two incompatible suffixes on verbs for instance. Rather, they choose one verb suffix, then, as if to add a further boundary between the chosen clause types, they add a lexical coordinator such as aka ‘then’. This multiple juncture-marking has the function of separating the clauses even further than the existing construction would have, and perhaps gives speaker and listeners time to parse the information. It seems that the use of aka in this way is that of a spacer, to separate clauses in what might otherwise be a rather dense onslaught of information.

There are other types of complexity as well. Speakers sometimes use chain-like structures in which subordinate and medial clauses are linked together, seemingly alternating. In the following examples, note the use of subordinate anterior clauses (with suffix -ge -ANT) interspersed with clause chains (with suffix -re -NFIN).

(28)  
\[
\text{[Le-ne e-sia-ge], o-tua la kini ui} \\
\text{day-PERL 3sn.OBJ-SBD-do-ANT 3s.POSS-wife(F) SG.F.DEF ACTN food(N)}
\]
\[
\text{ga e-kui-re], [e-kui-re] [o-e-lo-ge], hano.} \\
\text{SG.N.DEF 3sn.OBJ-burn-NFIN 3sn.OBJ-burn-NFIN 3sf.OBJ-SBD-finish-ANT then}
\]
‘The day having come, his wife cooking the food, having finished cooking it, okay.’
(29) **O-tua la kini [namu vo-ta-re]**
3s.POSS-wife(F) SG.F.DEF ACTN place(PL) 3p.OBJ-clear-NFIN

3p.OBJ-clear-NFIN 3sf.OBJ-SBD-finish-ANT 3sf.OBJ-tell
‘His wife clearing the places, her having finished clearing them, he told her.’

(30) **[O-e-lo-ge], [a-igu-re], suma na**
3sf.OBJ-SBD-finish-ANT 3sm.OBJ-take.out-NFIN taro(M) SG.M.DEF

o-kui-m na, aka / fofo tula o-na fi /
3s.SBJ-burn-SG.M SG.M.DEF then / basin(F) small 3sf.OBJ-in 3sn.FOC /

mina / maki na a-o-ho.
thing(F) / nut.pudding(M) SG.M.DEF 3sm.OBJ-3s.SBJ-put.inside
‘Having finished, taking it out, the taro she had cooked, then she put the nut
pudding into a small basin.’

(31) **[O-u-re] [lo-re], fi aka / a-na kini**
3sf.OBJ-eat-NFIN finish-NFIN 3sn.FOC then / 3sm.OBJ-in ACTN

koi [ruio la o-ne-ge],
also bamboo.water.container(F) SG.F.DEF 3sf.OBJ.3s.SBJ-give-ANT

o-laf ga e-u-o.
3s.POSS-water(N) SG.N.DEF 3sn.OBJ-eat-?
‘Finishing eating it, then her giving him a bamboo water container,
he drank his water.’

(32) ... **[dani ga e-ki-ge,] [e-le-re], [malav**
3sn.OBJ-SBD-dawn-ANT 3sn.OBJ-see-NFIN people(PL)

va ma-fan ika vo-lufu-re] o'ase-ne gala
PL.DEF 3p.POSS-some there 3p.OBJ-leave-NFIN bush-Perl there

kini vau ma-ao.
ACTN go.seawards 3p.SBJ-go.in
‘… dawn having come, (they) seeing it, they leave some of the people there,
[the rest] go in [the canoe].’

Chains like this can be translated as ‘Having X-ed, Y-ing, she Z-ed’. This complex
interaction between cosubordinate and subordinate chains is a preferred way of structuring
sequences of events in Lavukaleve. It is not a case of a messy mixture between discrete
structures; it is better to think of it as the result of being built up from a number of
relatively straightforward structures, which come together in a complex way to express
complex sequencing of information.

### 2.5 Tail-head linkage

Complex constructions are also used in constructions larger than single sentences, that
is in tail-head linkage. Tail-head linkage can be used to join simple clauses. In the
following, the verb ‘return’ of the first sentence is repeated as the first verb of a clause
chain in the second sentence.
In (34), *fou ‘put–on’ is repeated from the final verb of the first clause as the first verb of the following clause.

(34)  \[
\text{Hoia-ri-om} \quad o-rolo-n \; \text{ka} \\
\text{DEM.NTRL.G-F-PR-M/N} \quad 3s.\text{POSS-inside-LOC} \; \text{LOC.EMPH} \\
\text{fi} \quad e-o-fou. \\
3s.\text{FOC} \quad 3s.\text{OBJ}-3s.\text{SBJ}-\text{put.on} \\
\text{‘That one she put right inside.’} \\
E-fou-re, \quad \text{foiga} \quad \text{“E-} \text{ma-vele} \; \text{vela-ma!”} \\
3s.\text{OBJ}-\text{put.on-NFIN} \quad \text{PRON.NTRL.G-N} \quad 3s.\text{OBJ}-\text{take-SUCC} \; \text{go-DUR.IMP.SG} \\
\text{‘Putting it inside, then “When you’ve taken the coconut, go!”’} \\
\]

Some of the instances of subordinate clauses are in the context of tail-head linkage (about one in five). Providing the backdrop for the rest of the sentence means subordinate clauses lend themselves to tail-head linkage constructions, which themselves structure information into given and new. In the small corpus, about a third of all subordinate structures were involved in tail-head linkage. These constructions repeat the last verb of a sentence, creating a structure in which a new sentence begins with given material, taken from the previous sentence, leaving the rest of the sentence to express the new material. For example:

(35)  \[
\text{“} \text{Ta! Kua} \; \text{hoga} \; \text{lokosu} \; \text{foga} \; \text{ke!”} \quad / \\
\text{just coconut.tree(N)} \quad \text{DEM.PRXL.G-N} \; \text{headless PRON.PRXL.G-N EMPH} \quad / \\
o-re. \\
3s.\text{SBJ-say} \\
\text{‘“Hey! This coconut tree has no top!” he said.’} \\
A-e-re-ge, \quad \text{sina-n} \; \text{fi} \\
3s.\text{OBJ-SBD-say-ANT} \; \text{whisper-LOC} \; 3s.\text{FOC} \\
o-kelago-ri \quad \text{ma-me}. \\
3s.\text{OBJ-go.over-CAUS} \; 3p.\text{SBJ-HAB} \\
\text{‘Him having said this, they passed [the words] over [the line] in a whisper.’} \\
\]

(36)  \[
... \text{a-} \text{rive-re} \quad \text{kini} \; \text{o} \text{-feu}. \\
3s.\text{OBJ-follow-NFIN} \; \text{ACTN} \; 3s.\text{SBJ-go.inland} \\
\text{‘... she follows him inland.’} \\
\text{Kini} \; \text{o} \text{-feu-ge} \quad \text{ta} \; \text{oina} \quad \text{soi-m}. \\
\text{ACTN} \; 3s.\text{OBJ-SBD-go.inland-ANT} \; \text{just other.NTRL.M} \; \text{run.away-SG.M} \\
\text{‘Her having come inland, the boy runs away.’} \\
\]
3 Conclusion

At the outset it was hoped that it would be possible to add to the morphosyntactic description of Lavukaleve predicate or clause combining constructions by coming to some understanding of the separate functions of these constructions. To the extent that I have succeeded, I hope it is clear that the type of construction (or constructions) shown in the last sections are not messy aberrations. Rather, these structures follow principled patterns of information flow. Indeed, this is how much of Lavukaleve discourse is structured. By means such as have been described above, Lavukaleve speakers pace their information, using SVCs to unite multiple events into single chunks; coordinate constructions to link clauses in temporal sequences; cosubordinate constructions to separate information into different chunks of content or aspect, subordinate constructions to give a backdrop to the narrative; and tail-head linkage to provide coherence across larger boundaries.

References


~ Lexicography, formulaic language and the lexicon ~
When first invited to contribute to this celebration of Andrew Pawley’s academic life and works, I had intended to take what I thought was the easy road and simply contribute a few personal reminiscences. However when I actually sat down to do this, I realised that I would not be able to write more than a paragraph without implicating other people in a way that they might not appreciate. Andy has always been a compassionate and solicitous shepherd of his flock of graduate students, and the best stories I could tell would involve some of the more errant or befuddled among his disciples, who, in this age of Google, might still be inadvertently identified as the subjects of an anecdote. So, while I will mix a few blander reminiscences into this article, I will say no more about Andy’s interactions with his students, except that I can think of few people who could come close to matching Andy in his patience and unfailing concern for the least promising as well as the most brilliant of them; I am sure that these qualities have also been a hallmark of his relationships with his colleagues.

I first knew Andy when I was an undergraduate and he was a graduate student at the University of Auckland in the days when anthropology subsumed both linguistics and Māori studies, thus ensuring that students like Andy got a really good grounding in anthropology and also both anthropological and Bloomfieldian linguistics. Our initial academic journeys followed similar paths, as we each in turn studied at the University of Hawaii (although Andy’s degree was conferred by the University of Auckland), and married fellow students born in the Philippines whom we had met there. Andy had gone back to New Zealand by the time I left for Hawaii, but after I returned five years later we became close friends and associates, although Andy was based in Auckland (and for a while also commuted to Hawaii) while I was based in Wellington (with sporadic sojourns in Singapore and the Philippines). Our families have remained in pretty regular contact ever since, and I am delighted to be able to express my appreciation to Andy for his support and assistance in many matters, weighty and trivial, over the years. One of these matters which I think joins those extremes is my developing ‘Language Garden’, for which Andy has been a source of both encouragement and information; the title of this contribution reflects one aspect of that project.

Among his many significant contributions to intellectual endeavour, Andy’s work on historical and comparative linguistics has been especially important, and the three volumes so far published on the Proto Oceanic lexicon co-edited by Malcolm Ross, Andy, and
Meredith Osmond are in their own way as significant for Austronesian studies generally as Dempwolff’s iconic Vergleichende Lautlehre des Austronesischen Wortschatzes has been for succeeding generations of scholars. I have found the third volume, on plants, extraordinarily valuable for the language garden project, and am grateful to Andy for having allowed me access to drafts of the chapters I was particularly interested in before the book was published. My major reference guide for the origins of the plant names brought to Aotearoa New Zealand by the early Polynesian settlers had been the unpublished Pollex database, set up by the late Professor Bruce Biggs and continued after his death by Ross Clark. (It was Bruce who inspired Andy and me, along with many others, to pursue graduate studies in linguistics when this was a very new field in New Zealand in the 1960s.) Pollex, while being an invaluable resource for anyone seriously interested in the history of Polynesian languages, concentrates of necessity essentially on glosses rather than definitions, and thus while the Proto Polynesian reconstructions it proposes are formally sound, the ‘meanings’ generally are signposts to useful starting points for their further elucidation. Robert Blust (another of Andy’s and my contemporaries at the University of Hawaii) points out that for semantically complex terms such glosses ‘show a kind of shotgun approach that tries to capture the sense of a term in English semantic categories. The real meaning comes out in the more discursive glosses ...’ (2009:335).

I became acutely aware of this when working on the etymologies of terms relevant to Māori customary law, a task which required ideally the formulation of adequate definitions in English of the Māori terms and, where applicable, their Proto Polynesian and earlier antecedents. The Danish scholar J. Prytz Johansen (1954) had drawn attention to the interrelationship of the ideas conveyed by the terms, au [the first person pronoun, ‘the kinship I’], mana [a concept combining notions of psychic and spiritual force and vitality, recognised authority, influence and prestige, and thus also power and the ability to control people and events], and tupu [growth, increase, development]. It seemed to me also that in some way there was a similar kind of complementarity between the ideas conveyed by mauri [‘the essence which gives a thing its specific natural character’] and tupu, and that this reflected notions that were widely held in the Austronesian world, and probably dispersed in the very early Malayo-Polynesian migrations. My research on cognates in other Austronesian languages of the latter terms strengthened this idea. To gather more evidence in support of my hypothesis, which I thought might constitute an original contribution to knowledge, I asked Bob Blust for access to the files of his Austronesian database, to which he readily agreed. Unsurprisingly, since he had studied the same data in even greater depth, I found that Bob had already made a note of this (some ten years ahead of me) in an excellent postscript to his collection of reflexes of Dempwolff’s Proto Austronesian *qudip:

Dempwolff (1938) reconstructed */qudip/ ‘to live’, and although this semantic reconstruction is justified, it appears incomplete in a number of respects due to differences in the “conceptual focus” (Blust 1991) of the German/English and Austronesian terms. Whereas English ‘living, alive’ denotes merely the presence of life processes, PMP */qudip/ evidently denoted a dominance of vitality as manifested in growth, flourishing, and being healthy, fresh (of plants), or green (of plants, wood). Reflexes in MP languages show recurrent references to vegetation and to growth, a component of the meaning which is reinforced by the observation that */qudip/ has been replaced in a number of the languages of Sulawesi by reflexes of */tumbuq/ ‘to grow’. If anything in English reminds us of this conceptual focus it is perhaps the depiction of the life-force in Dylan Thomas, ‘The force that through the green fuse drives the flower’ (‘Drives my green age ... ’). Extensions of
this primary sense are seen in recurrent references to newness (Ngadha and ‘Are’are), the new moon (Roti, Maori, Hawaiian), and birth (Tetun, Sa’a) and perhaps the soul, where the vital principle is seen as essentially supernatural (Javanese, Lonwolwol, Rennellese, Anuta, Rarotongan, Nukuoro), and so applied to body parts that have well-known associations with the exit and entry of the soul (Rennellese: skull, fontanel, Hawaiian: fontanel). (Blust 1998, emphasis added)

For the Māori reflex of *qudip, mauri, Dylan Thomas’ lines quoted above encapsulate this very complex concept better than any set of shotgun glosses; similarly, tupu, the reflex of PAN *tumbuq, conveys in Māori the notion of growth and development, dynamism in contrast to regression and entropy; in a philosophical sense, ‘Tupu is man’s natural unfolding, which...denotes strength, courage, and honour’ (Johansen 1954:93). As the Sulawesi cognates confirm, it is not only in Māori that ‘growth’ and ‘increase’ have deeper, more metaphysical implications. Casting a wide net and using the evidence of historical and comparative linguistics, along with archaeological and contemporary evidence from other sources, has not only enabled Andrew Pawley, along with his collaborators at The Australian National University and other scholars like Robert Blust and Paul Geraghty to reconstruct the form and wider meanings of the sources of inherited words in contemporary languages, but also to throw considerable light on migration patterns, the environment, and other aspects of the history and experience of Austronesian peoples at various stages of their collective and separate development.

One outcome of this, inspired partly by conversations on linguistic matters with Andy Pawley and his admiration of my family’s somewhat anarchic but flourishing transformation of our property in Ngaruawahia, New Zealand, has been my exploration of the mauri and tupu of the names brought with them and bestowed on local plants by the Polynesian settlers of Aotearoa, along with an ongoing attempt to collect and grow as many of them as will flourish in our local environment. Paul Geraghty has recently (2009) reviewed most of the relevant Proto Polynesian and Proto Eastern Polynesian plant names in his discussion of the lexical evidence for the nature of the Eastern Polynesian homeland. Bruce Biggs (1991) who had earlier identified 85 protowords applied to a similar number of New Zealand plant species (with some names designating more than one plant, and some plants having several inherited names), discussed the possible reasons for the way these names had been used in their new environment, and explored the commonalities that might explain the sometimes puzzling use of the same inherited word (alone or in combination with other elements) to refer to several often quite disparate species. (One example, the Māori reflex of PPN *kafika, is discussed below.) Since Professor Biggs compiled his list of inherited plant names present in Māori, more have been added to the Pollex database, and recent advances in information technology have made it possible to add considerably to the number of species covered in part or wholly by these names. The Māra Reo (Language Garden) database (Benton 2009) for example currently includes over 130 protoforms referring to about 150 distinct plant species.

The authors of The Lexicon of Proto Oceanic, Volume 3, have identified a number of terms which refer to a variety of species as ‘taxa’, i.e. terms denoting a set of plants with certain characteristics, rather than regarding them as the names of individual plants. A plant name may also function as a ‘taxon’ in a very subtle way, that is, its wider taxonomic properties may not be apparent in the way it is normally used in any one language, but become apparent when it is applied in new environments. Some manifest such characteristics from the outset, and others develop wider taxonomic properties in places more distant from their point of origin. Although guesses can be made on the basis of
second-hand evidence, such as dictionaries, botanical treatises, and so on, there is a great advantage in the etymologist’s being personally acquainted with the plants as well as the names. One of the huge strengths of the plant volume in the Lexicon of Proto Oceanic series is that the editors and authors have been able to combine such personal knowledge with their scholarship. My own understanding of the antecedents of the inherited plant names in the New Zealand flora was greatly enhanced by the three months I spent in Hawaii in 2007, looking for the Hawaiian plants whose names had cognates in New Zealand Māori (and finding representatives bearing all except one of the names recorded in Pollex, including a number of species which shared these names but had not been included in the glosses in Pollex or other widely available comparative lists).

A good example of such a hidden taxon is, I believe, the Proto Eastern Oceanic and Proto Polynesian term *milo, glossed probably wherever it is mentioned as ‘Thespesia populnea’. This is most certainly the surface meaning of the word, but I was alerted in Hawaii to the wider meanings of the word, which I think were probably present, albeit latently, in Proto Polynesian at least. My interest was in discovering why in Māori the reflex of this name is designated Prumnopitys ferruginea, a podocarp and thus about as distant botanically from Thespesia as a tree can get. In Easter Island, the word miro is used as a taxon for timber or ornamental trees — trees whose fruit is not eaten by people. Furthermore, a Proto Eastern Polynesian form *toromiro also incorporates this word. In (New Zealand) Māori this word refers again to Prumnopitys ferruginea, and in Tahiti to the Thespesia in its function as a sacred tree, but in Easter Island it denotes a leguminous tree (Sophora toromiro) very closely allied to the New Zealand kōwhai (S. tetraptera), and in Rarotongan Māori to another leguminous tree, Schleinitzia insularum.

Although very different superficially and botanically, these milo/miros all have a lot in common: they have very strong, beautifully grained and easily worked wood, visually attractive fruits which are not food for people, and an aesthetically-pleasing form. Miro has always been one of my favourite trees because of its gracefulness and the way its branches move in the wind, and when I first saw the Hawaiian milo I wondered why the Polynesian settlers of Aotearoa had used this name for an apparently quite different tree. A probable explanation was provided in a book on Hawaiian flora which formed part of the exhibits at the Kōhee Museum on the island of Kaua‘i:

The leaves are beautifully glossy, and the wind moves them in a most graceful way, somewhat like the quivering of the aspen. (Sinclair 1885:10)

A similar appreciation of the milo’s aesthetic qualities was expressed independently by a botanist friend. The name *milo may thus have carried far more than a just reference to a particular species; rather it appears to have evoked a number of special qualities that link several quite different trees along the route which the early Polynesian explorers travelled. There is some debate as to whether this tree is truly indigenous to Hawaii, or was brought there from Tahiti or the Marquesas, along with taro, aute (paper mulberry) and other plants by the Polynesian settlers. Like the karaka in New Zealand (which is an indigenous plant, and about which more below), it is associated with human settlements especially around the coasts, which is not surprising considering the esteem in which the wood and the tree itself were held.

In Tahiti, the miro (Thespesia) is a sacred tree, referred to in this context by a homophonous reflex of Proto Eastern Polynesian *toromiro. As noted above, in Tahitian and New Zealand Māori the modern reflexes of *toromiro refer to the same tree as those of the older form *milo, from which it is obviously derived, but in both Rapanui and
Rarotonga they refer to what seem to be quite different trees. However, there are links among these divergent meanings. In Tahitian, and probably also in New Zealand Māori (I have not come across any examples of the use of the word *toromiro apart from the minimalist entry in the Williams dictionary) *toromiro refers to a special aspect of the miro, its sacred quality. There is a useful early comment on the Tahitian *amae (toromiro) tree in the journal of James Morrison, the Boatswain’s mate of the Bounty, published under the title ‘Account of the Island of Tahiti and the Customs of the Island’, and held in the Australian National Library:

> Amai — this is a hard Close Wood of a reddish ColLOUR which they also used in Building their Canoes, making Ax & adze handles &c. — with this tree they Generally plant their Morai’s [marae] or places of Worship, & the leaves of this tree are always used in religious rites.

In Rapanui and Rarotonga the newer word refers to leguminous trees. The Rapanui one at least has many of the qualities attributed to the ‘original’ *milo (Thespesia polulnea) — a beautiful tree, with greyish bark, and a very hard, beautifully-grained reddish wood, qualities which link all these trees to the New Zealand miro, belonging to a very different family but also highly significant aesthetically and culturally. (The New Zealand miro was also economically important as a prime food source for birds, which could then be caught and preserved for eating out of season).

Bruce Biggs discusses several similar instances in his 1991 article, but treats them more as on-the-spot analogies rather than manifestations of more fundamental latent semantic associations in the inherited word. He notes the absence of the coconut palm in New Zealand, and lists many of the terms referring to parts of the coconut, most of which are not now reflected in Māori. Two are, however: *nīkau ‘a coconut frond’ becomes *nīkau ‘the palm Rhopalostylis sapida’, and the word *niu itself, which came to mean a ‘divining stick’. These are listed in the appendix but not commented on in Prof. Biggs’ paper. The part to whole transformation, as with *nīkau, is clear enough — the *nīkau has no coconuts, but it is definitely a palm with very similar fronds. The fate of *niu is more interesting. First it becomes a divining stick (a miniature coconut trunk), then, in the mid-nineteenth century, it regains some of its height and girth, being used by the Hauhau devotees as the term for the tall ceremonial poles they set up at their sacred sites. However, not enough of its original meaning remained for the word to be revived for its erstwhile referent, as in modern Māori the English-derived kokonati is used for both the nut and the palm which bears it.

Another set of words discussed by Bruce Biggs are the reflexes of Proto Polynesian *kafika. Like *milo, this term seems to have had a specific referent in Proto Polynesian times, in its case the Malay apple, Syzygium malaccensis. However, both the word and its meaning have undergone some interesting transformations in Polynesia, as can be seen from the list of reflexes (mostly from Pollex) below. There is a very interesting convergence between Hawaiian and Māori, with the application of a variant of the name to the species of rātā (Metrosideros) — not difficult to explain when you compare the Syzygium flowers with those of the Metrosideros species.

- Tongan: feikika (Syzygium malaccensis) [Myrtaceae]
- Niuean: feakai (S. malaccensis) and Kafika (S. inophylloides)
- Samoan: nonu fi’afi’a (S. malaccensis)
- Tahitian: ‘ahi’a (S. malaccensis)
Marquesan: kehika (*S. malaccensis*)  
Hawaiian: ‘ōhi‘a ‘ai (*S. malaccensis*) and ‘ōhi‘a, ‘ōhi‘a lehua  
(*Metrosideros* spp.) [Myrtaceae]  
Rarotongan: ka‘ika (*S. malaccensis*)  
Māori: kahika, kahikatea (*Dacryocarpus dacrydioides*) [Podocarpaceae], kahika, flower of the rātā (*Metrosideros robusta*), kahikahika (*Metrosideros fulgens*), and kahikaatoa (*Leptospermum scoparium*) [Myrtaceae]

The kahikaatoa (kahika-a-toa) ‘warrior kahika’ is related botanically to the *Syzgium* (they belong to the myrtle family, *Myrtaceae*) — its fruits are certainly not edible, but the unopened buds (as well as the durable timber) are reminiscent of the *kafika*, and the centres of the flowers are most certainly reminiscent of the rātā and pōhutukawa, which also belong to the *Myrtaceae*. How the kahikatea (‘white kahika’) fits into this equation is slightly more problematic, yet it has also inherited the name without any qualifications (kahika), so a link must be there! My guess is that it would be the slim, straight trunk which would have attracted instant attention, and probably merited the name when the Polynesian *Syzygium* was nowhere to be found. The small orange berries festooning the tree in summer, although quite different in form, might have reinforced the allusion.

The kahikaatoa, however, has an alternative name, mānuka, which it shares with a related species, *Kunzea ericoides*. Although the Pollex database does not do so, I would link this name with both Proto Fijic (Proto Central Pacific?) *nuka* (referring, as a plant name, specifically to the tree *Decaspermum fruticosum*), and also, more remotely and indirectly, with Proto Austronesian *nukaq* ‘wound’. The Pollex database has three reconstructions encompassing this set of plant names:

*NUKA-NUKA (ultimately Proto Fijic) ‘A shrub (*Decaspermum* sp.)’  
Fijian: nuqa, nuqanuqa (*Decaspermum fruticosum*) [Myrtaceae]  
Tongan: nukonuka (*Decaspermum fruticosum*)  
Samoa: nu’anu’a (*Decaspermum fruticosum*)  
East Uvean and East Futunan: nuqanuqa (*Decaspermum fruticosum*)  
Tahitian: nuanua (*Decaspermum fruticosum*);  
(Pollex also gives *Nelitis vitiensis* as a gloss for Fijian *nuqunuqa*, but I cannot find references elsewhere to a plant of this name.)

*MAA-NUKA (ultimately Proto Austronesian, from *nukaq ‘wound’) ‘Tree sp.’  
Rennellese: manguka (*Altonia spectabilis*) [Apocynaceae]  
Rarotongan: manuka (*Leptospermum scoparium*)  
Māori: mānuka (*Leptospermum scoparium* and *Kunzea ericoides*) [Myrtaceae]);  
mānuka rauriki (*Leucopogon fasciculata* [Ecpridaceae] and *Kunzea ericoides*)

*KAAKU (Proto Tahitic)  
Tahitian: anua (‘A tree with strong wood’)  
Māori: kānuka (*Kunzea ericoides*)

It would seem to me that all three of these reconstructions incorporate metaphorically Proto Austronesian *nukaq, in the sense that the various tree species all have either medicinal properties, thus are potentially remedies for wounds, or hard wood suitable for
making spears and other weapons to inflict wounds, or both. I am not sure whether the Renellese term really is cognate with the Māori and Rarotongan forms. I have not been able to find any specific information about the plant to which the Renellese name apparently refers, but the closely related *Alionia scholaris*, which grows in Southeast Asia and parts of the Pacific is a tall forest tree noted in Indonesia and Papua New Guinea for its medicinal properties. If *manguka* is cognate with *mānuka*, then *m(a)nuka* would clearly be Proto Polynesian; if not, it would derive from the period before the two Tahitian languages separated. *Kānuka* seems certainly to have originated in Tahiti. Both, however, are related to a form I would reconstruct as *(n)uka*, from the period before the Polynesian and Fijian languages separated. What I imagine to be the original ‘nuka’ is featured on a Fijian stamp, with this description in the Fiji postal service’s publicity material written by Paul Geraghty:

> The nuqanuqa (also widely known as nuqa, and niqwa in parts of Westen Fiji) is a moderate sized tree found at the edge of the forest, particularly in dry, rocky places, and also cultivated. It is endemic to Fiji, and has small round pointed leaves which give off a pleasant smell when crushed, small black berries, and small white fragrant flowers. The nuqanuqa is also medicinal; an infusion of its leaves, with the leaves of certain other plants, is said to cure diabetes, the bark makes a poultice for piles, and an infusion of its root, with root of vobo, is said to cure cervical and breast cancer. (Post Fiji 2006)

Of the New Zealand plants that have inherited these names, only the alternative name for the *mingimingi* (*Leucopogon fasciculata*) is a little puzzling, in that *mānuka rauriki* literally means ‘small-leaved *mānuka*’ — its leaves are certainly not small in relation to the other New Zealand *mānukas*. However, the reference could be directed back to Polynesia, ironic, or motivated by something else. There may well be something in traditional accounts or poetry which will shed some light on this mystery. Since there are no *Decaspermum* in the New Zealand flora, the other *mānuka*, with their twin healing and wounding potential, seem good candidates to inherit these names.

There is at least one very important New Zealand tree whose name is certainly a case of mistaken identity. The *karaka*, *Corynocarpus laevigata*, bears a name derived from Proto Oceanic *(k)alaka* ‘*Planchonella* sp.’, with a similar form and referent in Proto Central Pacific. The Fijian reflex, *qalaka*, refers to *Pouteria (Plancholella) costata*, a species which is also present in New Zealand. (There is considerable debate among botanists about the membership and status of the genera *Pouteria* and *Planchonella* within the family *Sapotaceae*, and this species has been shifted from one to the other several times in recent decades.) In Tongan, Samoan, and the Eastern Polynesian languages, except for New Zealand Māori, the cognate words refer mostly to a *Planchonella/Pouteria* species:

- **Tongan**: kalaka. A tree (*Planchonella grayana*)
- **Samoan**: ala’a. A tree (*Planchonella sp.*), a shrub (*Charissa sp.*)
- **Hawaiian**: ’ala’a. (*Planchonella sandwicensis*)
- **Tuamotuan**: karaka. Stimpson records this as a tree resembling the *rama*, and the *rama* as a shrub with ‘a bright round orange-coloured fruit about the size of a small egg, which contains edible flesh and a kernel’
- **Rarotongan**: karaka (*Elaeocarpus rarotongensis* [*Elaeocarpaceae*] and *Planchonella grayana*)
- **Māori**: karaka (*Corynocarpus laevigata* [*Corynocarpaceae*])
A feature of these trees is their attractive, bright green leaves and relatively large, brightly coloured berries. It is thus easy to understand how someone who was not a botanist could have mistaken the New Zealand tree Corynocarpus laevigata for the Polynesian karaka, and applied the name before the false identity was discovered. It certainly would have been revealed the first time anyone tried to eat the fruit and bit into the highly toxic kernel (when ripe, the flesh is quite safe to eat). Dealing with those fruit would not, however, have been a great problem. The kernels became an important staple food item, detoxified and made palatable in pretty much the same way that the Tahitian chestnut (Inocarpus fagifer) is prepared in tropical Eastern Polynesia, a procedure with which the early Polynesian settlers were no doubt very familiar. The Corynocarpus karaka is endemic to a small part of Northern New Zealand, around the Bay of Islands, but was carried from there throughout Aotearoa and also to the Chatham Islands, undoubtedly giving rise to the tradition that the tree was brought to New Zealand from Hawaiiki. Meanwhile, the ‘real’ karaka, Pouteria costata, remained hidden in the coastal forest and was later given the local name of tawapou.

Legend can also play an important part in the naming and classification of plants. In some Māori typologies southern right whales (tohorā) and the kauri tree (Agathis australis) are closely linked, because of the similar appearance of their skins — according to one legend, they were good friends, but because the kauri preferred the land to the sea, the whale suggested they exchange skins, and thus the kauri has a thin rough bark and resin to parallel the whale’s oil. In the light of another link between creatures of the sea and the land, I will conclude this very brief account of aspects of the secret life of plant names with a proposal that a Hawaiian plant name be added to the list of reflexes of the Proto Polynesian *fāpuku ‘grouper’ (Epinephelus sp.). Reflexes of this word are widespread in Polynesian languages, and in most refer to one or more of several related species of grouper. This is the case also in Hawaiian, where the word hāp‘u refers to the Hawaiian grouper, E. quernus. However, when I was in Hawaii it seemed to me highly improbable that it was a mere coincidence that this was also the name of the four species of tree ferns belonging to the genus Cibotium: C. chamissoi and C. nealii (hāp‘u), C. glaucum (hāp‘u pulu) and C. menziesii (hāp‘u i‘i‘i), but I could not find any solid evidence to support this hypothesis. The solution to this puzzle appeared when I was checking out the whale story while writing this paper. On the Encyclopaedia of New Zealand web site, it is noted that ‘Another tradition cites Te Hāpuku [rather than Tangaroa or some other contenders] as the main ancestor of whales, dolphins and seals as well as tree ferns, which are often known as ‘ngā ika ō te ngahere’ — the fish of the forest’. All that remains now is to find a Hawaiian version of this tale.

During his long and continuing career as a scholar Andrew Pawley has emulated Te Hāpuku by producing notable works on a range of topics and languages as disparate as dolphins and tree ferns. Along with the other contributors to this volume, I wish him ‘Ad multos annos’, and look forward to further fruits of his labours, including more volumes of the Proto Oceanic lexicon.

References


Five patterns of semantic change in Austronesian languages

ROBERT BLUST

1 Introduction

The primary purpose of this paper is to draw attention to certain recurrent patterns of semantic change in Austronesian languages. A secondary, and to some extent unavoidable purpose, is to relate these patterns to the broader literature on semantic transfer, both synchronic and diachronic. In doing this it becomes apparent that the literature on diachronic semantics is inadequate in various respects. While it can force most — perhaps even all — semantic changes into a procrustean bed of metonymy, metaphor and synecdoche, it does so at the cost of losing sight of shared properties that are not captured by these labels, or that are captured only in an indirect and fundamentally awkward manner. To address this problem I have proposed some alternative ways to classify types of semantic change. These alternatives are not intended as replacements for traditional or current typologies, but rather as an indication that other types of classification are possible which are motivated by different research questions, and so give rise to different types of insights into the diachronic study of meaning. To place my proposals in proper perspective I will first briefly review the history of scholarship concerning semantic change.

1.1 Brief history of research on semantic change

In this section I do not pretend to present a detailed history of research on semantic change. Rather, my goal is to touch on a few highlights in this field of inquiry so as to give

---

1 I offer this paper as an inadequate expression of my admiration for the honoree of this volume. Andy Pawley is many things to many people: methodologist par excellence, pioneer in Oceanic comparative grammar and the still nascent world of Papuan historical linguistics, major contributor to Polynesian and Fijian descriptive and comparative linguistics and to English conversational analysis, one bold enough to take on a language ‘which defies description by ordinary means’ (Pawley 1993), father of festschrifths — at least five, including blockbusters for an ethnobiologist (Pawley 1991), and a Pacific archaeologist (Davidson, Irwin, Leach, Pawley and Brown 1996), reflecting the wide range of his academic pursuits and professional relationships, editor, administrator, author at age 20, walking encyclopedia of sports statistics. He is all that and more. For me he has simply been one of the most steadfast friends and colleagues I have had in the 35 years since I was lucky enough to take a graduate course from him during his time in Hawaii as a visiting scholar. If it were not for his unrelenting commonsense I probably would still be writing my doctoral dissertation (now approaching 20,000 pages).
a sense of how semantic changes have been classified and explained over time, and how interconnected (or otherwise) major lines of research have been.

1.2 Bréal

Although speculation about semantics goes back at least as far as classical antiquity, the earliest treatment of semantic change that is generally recognised is Bréal (1897), a book that was first published in 1884, and then went through five editions. In Part II, ‘How the meaning of words is determined’, Bréal examines how words ‘once created and endowed with a certain meaning, are induced to restrict, to extend, to transfer this meaning from one order of ideas to another, to raise or to lower its dignity, in short to change it.’ He identifies the following types/patterns of semantic change:

1. pejorative tendency
2. meliorative tendency
3. restriction of meaning
4. expansion of meaning
5. metaphor
6. concretion of meaning (‘an abstract word, instead of keeping its abstract sense, instead of remaining the exponent of an action, a quality, or a state, becomes the name of a material object’)

Although thorough and insightful for its time, this work now has only historical interest, its typology having been completely incorporated in or superseded by later works, as Bloomfield (1933).

1.3 Bloomfield

In his 1933 textbook, Language, Bloomfield devoted a 19-page chapter to ‘Semantic change’. In it he summarised the work of earlier scholars in recognising the following types/patterns of innovation; these largely incorporate the distinctions recognised by Bréal, with the exception of ‘concretion of meaning’, which is omitted:

1. narrowing: Old English mete ‘food’ > meat ‘edible flesh’
2. widening: Middle English bridde ‘nestling’ > bird
3. metaphor: Proto Germanic *bitraz ‘biting’ > bitter
4. metonymy – the meanings are near each other in space or time: Old English cēace ‘jaw’ > cheek, Old French joue ‘cheek’ > jaw
5. synecdoche – the meanings are related as whole and part: Proto Germanic *tu:nae ‘fence’ > town
6. hyperbole – from stronger to weaker meaning: pre-French *ex-tonāre ‘strike with thunder’ > French étomner ‘astonish’
7. litotes – from weaker to stronger meaning: pre-English *kwalljan ‘to torment’ > Old English cwellan ‘to kill’
8. degeneration: Old English cnafa ‘boy, servant’ > knave
9. elevation: Old English cniht ‘boy, servant’ > knight
1.4 Ullmann

In a book which summarised most of what was then known about the linguistic study of meaning Ullmann (1962) included an 87-page chapter on ‘historical semantics’. At the outset he states in somewhat Saussurean terms that ‘meaning’ and ‘semantic relationship’ are synchronic notions that are meaningless in a diachronic context, and that ‘semantic change’ is a diachronic concept that cannot be translated into synchronic language. Nonetheless, the two are seen as ‘interdependent, one being the projection of the other on a different plane.’ Ullmann (1962:183ff.) identifies the following influences responsible for semantic change:

1. pejorative tendencies: *hussy* ‘housewife’ > ‘woman of worthless character’
2. evolution of the referent: *papyrus* to *paper*, name fixed, but referent undergoes technological improvement
3. metaphor motivated by a gap in the vocabulary, or by artistic delight: ‘sky-scraper’ coined to represent new referent, ‘a galaxy of beauties’ to express delight
4. taboo: circumlocutions of a euphemistic and propitiatory nature for names of dangerous animals such as the bear in Europe
5. euphemism: French *tricher* ‘cheat, trick; doctor’
6. infection: as a result of syntagmatic contiguity the French negative particle *ne* has ‘infected’ Latin *persona*, causing it to become a negative particle itself
7. passage into a specialised social group: English *bishop* and French *évêque* originally meant ‘overseer’, but acquired a religious sense as a result of passing into a specialised social group
8. rendering abstract processes more tangible: English ‘to grasp’, meaning ‘understand’, derives from the more concrete action of clutching in the hand
9. semantic imitation: some languages may imitate the semantic changes of others. Thus, German *Schöpfung* ‘act of creation’ acquired in the 18th century the additional sense of ‘all created things’ by imitation of the ‘nomen actionis — nomen acti’ polysemy of English ‘creation’.

Only two of the factors that Ullmann implicates in semantic change are mentioned by Bloomfield (pejorative tendencies, metaphor). The comparison of these lists suggests that until at least the early to mid 1960s there was no general agreement about how to develop a theoretically insightful typology of semantic change. In the several decades since the appearance of Ullmann’s book no new typologies have been proposed, but in their place there has been a renewed interest in defining the difference between certain terms that were used rather loosely in earlier studies of semantic change, most notably the distinction between metaphor and metonymy on the one hand, and metonymy and synecdoche on the other.

1.5 The cognitive linguistics paradigm

Over the past quarter century there has been an explosive increase in scholarship on semantic change, together with a corresponding increase in theoretical sophistication. Some of this has little bearing on the claims made here, as Traugott and Dasher (2002),
which is concerned primarily if not exclusively with the semantic history of grammatical morphemes. The work that is most relevant to what I have to say in this paper has been produced within the field of cognitive linguistics, a subdiscipline that has roots in the seminal work of Lakoff and Johnson (1980). Volumes that contain important studies within this tradition include Blank and Koch (1999), Panther and Radden (1999), Dirven and Pörings (2002), and Panther and Thornburg (2003). A volume that lies outside this tradition, but which is valuable as a major collection of relatively recent papers on the topic of semantic change is Fisiak (1985).

In discussing semantic change in Austronesian languages reference will be made to specific papers in some of these volumes. One general comment that is worth making now, however, is that the study of semantic change by linguists from Bréal on is almost entirely restricted to Indo-European material, and any claims to universality are therefore tainted by the usual problems attaching to biased data samples. That this is a valid and important criticism has recently been made clear by Evans and Wilkins (2000), who have shown that earlier claims about the universality of patterns of polysemy and semantic extension in the domain of perception verbs do not apply to Australian languages (in particular, the verb ‘to hear’ rather than the expected verb ‘to see’ is the basis for semantic extensions to verbs meaning ‘to think’ and ‘to know’).

1.6 The anthropological tradition

The brief history sketched so far is limited to work on semantic change by linguists. It would be incomplete, however, if it did not include work on semantic change done within the field of cultural anthropology. Just over a decade before Lakoff and Johnson published their seminal work on metaphor, Berlin and Kay (1969) published an equally seminal work on the patterning of colour terminologies in the world’s languages, and the implications of this patterning for schemas of semantic evolution. Working within this framework Cecil Brown published a number of studies, most notably Brown and Witkowski (1981a, b), Brown (1983) and Brown (1984), in which he developed the notion of ‘universal change paths’, or recurrent paths of semantic evolution in the world’s languages. A salient feature of these studies is the broad empirical base used to support inferences, a striking contrast to most work on semantic change within the linguistic tradition. To cite just two representative examples, Brown and Witkowski (1981b) document a ‘pupil of the eye’ = ‘child/person of the eye’ equation in 25 languages representing at least six language families, and a ‘biceps/calf’ = ‘mouse/toad/lizard of the upper arm/lower leg’ equation in 23 languages representing at least five language families. These symbolic equations are interpreted as synchronic residues of historical changes in which speakers recurrently recruit members of these body-part categories from similar types of lexical sources as a result of cognitive processes common to all human languages.

---

2 Evans and Wilkins (2000:547) mention Sweetser (1990), who claims ‘predominantly on the basis of Indo-European data’ that a vision-based vocabulary relating to thinking and knowing is universal. Other studies that Evans and Wilkins note avoid sample bias, but are still shown to be overgeneralising. All of Bréal’s and Ullmann’s examples appear to be Indo-European, although they represent various branches of the family. Bloomfield’s illustrative material is drawn entirely from West Germanic languages and French. Another work published at about the same time (Stern 1931) is richly illustrated with examples, but draws its material almost entirely from English. One of the few exceptions to this Eurocentric bias is Canart (1979), who cites Indo-European, Japanese, Chinese and Malay data, but is concerned only marginally with diachronic semantics in the sense that this is understood by historical linguists.
1.7 Overview

What is initially surprising about these traditions is their insularity. Panther and Radden (1999), for example, contains eighteen papers, and none of these refers to any publication in the anthropological tradition mentioned above. Similarly, Blank and Koch (1999), which is specifically concerned with diachronic semantics, contains thirteen papers, and the sole reference in any of these is to Berlin and Kay (1969) in a paper by John R. Taylor. This insularity, however, reflects the nature of scientific paradigms: as noted decades ago by Kuhn (1962), paradigms are defined by the questions they address, and different fundamental questions lead research in different directions. Whereas writers like Berlin and Kay, or Brown and Witkowski are concerned with developing typologies of implicational universals and evolutionary schemas that presumably give rise to them, writers within the current cognitive linguistics tradition are far more preoccupied with clarifying the conceptual basis of their field. One entire volume (Dirven and Pörings 2002), for example, is devoted to a comparison and contrast of metaphor and metonymy in an effort to reduce problems that have been encountered in distinguishing these mechanisms of semantic transfer, and a number of papers in other volumes (as Panther and Radden 1999, or Panther and Thornburg 2003) have a similar aim. The general consensus of this line of research is that metaphor has long been the focus of attention at the expense of metonymy, a tendency that is essentially reversed in more recent work, where metonymy is seen to be perhaps the central mechanism of semantic transfer (Pauwels 1999). Before proceeding it will be best to briefly summarise recent conclusions regarding the distinctions made between metaphor, metonymy and synecdoche, since in the context of the current dominant paradigm (cognitive linguistics) these will be seen by some readers as issues in relation to the data I discuss.

Lakoff and Johnson (1980:36) hold that ‘Metaphor is principally a way of conceiving one thing in terms of another, and its primary function is understanding. Metonymy, on the other hand, has principally a referential function, that is, it allows us to use one entity to stand for another.’ In illustration, the sentence ‘Inflation robbed me of my savings’ is given as an example of metaphor, since ‘inflation’ is not being used to refer to a person, but rather to an abstract process which has consequences that are in some sense parallel to one being robbed by a human being. By contrast, the sentence ‘The ham sandwich is waiting for his check’ does designate a person (who ordered a ham sandwich), and thus serves a referential function rather than exploiting symbolic parallelism to achieve greater expressive power. They further conclude that synecdoche, where the part stands for the whole, is a special case of metonymy.

More recent treatments suggest that the distinction between metaphor and metonymy is not always clearly drawn, or that Lakoff and Johnson erred in calling synecdoche a type of metonymy. Koch (1999:139), for example, notes that ‘metaphor and metonymy … often get confused, not only by students, but — as to the application of these two theoretical notions — even by scholars.’ In a particularly useful paper Seto (1999:91) points out that ‘Despite the amount of interest shown in metonymy during the past two decades, no precise definition of metonymy has been offered yet; most definitions are so vague that they might also cover other concepts such as synecdoche, metaphor, irony, etc.’. To at least partially remedy this situation Seto defines metonymy as ‘a referential transfer phenomenon based on the spatio-temporal contiguity as conceived by the speaker between an entity and another in the (real) world’ and synecdoche as ‘a conceptual transfer phenomenon based on the semantic inclusion between a more comprehensive and a less
comprehensive category.’ More succinctly, metonymy is called ‘an E(ntity)-related transfer’, and synecdoche ‘a C(ategory)-related transfer.’ In illustration, the arm is a part of the body, and so has an E-relation to it, or one of partonomy, while a fir is a kind of tree, and so has a C-relation to it, or one of taxonomy. Both metonymy and synecdoche can be diagrammed as a part-to-whole relationship, but this representation is ambiguous, since one is a ‘part of’ relationship and the other a ‘kind of’ relationship. Under this interpretation traditional examples of synecdoche, such as ‘three sails upon the horizon’ (for three ships), would be seen as examples of metonymy.

Much more obviously could be said, but would be of only tangential relevance to the main interest of this paper, which is concerned less with clarifying the basis between traditional categories of semantic transfer such as metaphor, metonymy and synecdoche, than with showing that these categories sometimes fail to capture interesting generalisations that can be expressed by other means. With this as background information I turn now to the primary topic of my paper, returning to the generalist literature only where this is needed to elaborate on the discussion of particular problems.

2 Patterns of semantic change in Austronesian languages

This section illustrates five patterns of semantic change in Austronesian languages: (1) categorial linkage and decoupling, (2) semantic fragmentation, (3) semantic change due to change of environment, (4) prototype/category interchange, and (5) symbolic avoidance. In some cases similar patterns are attested in Indo-European languages as well.

2.1 Categorial linkage and decoupling

Bloomfield (1933:430) suggested that many semantic changes follow a pattern which can be described by the following categorial progression.

\[
A \rightarrow A + B \rightarrow B
\]

He illustrated this with a three-stage diagram in which \textit{food} (A) initially meant ‘nourishment (1)’, \textit{meat} (B) meant ‘edible thing (2)’, and \textit{flesh} (C) meant both ‘edible part of animal body (3)’ and ‘muscular part of animal body (4)’. In Stage 2 \textit{meat} expanded its reference to include ‘edible part of animal body’, but not ‘muscular part of animal body’, which was still designated by \textit{flesh}. In Stage 3 \textit{food} expanded its reference to include ‘edible thing’. Schematically:\footnote{Bloomfield (1933:430ff.) calls this ‘expansion and obsolescence’ of meaning, and credits the Neogrammarian Hermann Paul (1880) with being the first scholar to recognise its significance.}

<table>
<thead>
<tr>
<th>Meaning</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

\textbf{Figure 1:} Bloomfield’s ‘expansion and obsolescence’ of meaning
Bloomfield’s example, which makes use of four categories of meaning, appears poorly suited to the point he is making, namely that a linguistic form that has a single well-defined meaning acquires a secondary meaning in association with the first, and then loses its original sense, leaving the secondary meaning as the only sense of the term. Where he infers an expansion of reference it is the virtually vacuous claim that *meat* moved from meaning ‘edible thing’ to meaning both ‘edible thing’ and ‘edible part of animal body’ (which presumably would be included under ‘edible thing’), and thence exclusively to ‘edible part of animal body’. Far better examples of the structural process $A \rightarrow A + B \rightarrow B$ are available, and unlike Bloomfield’s example, nearly all of these involve semantic changes that yield important inferences about culture history.

Three well-known examples from the history of English are *bead*, *pen* and *clock*. English *bead* is cognate with German *Gebet* ‘prayer’, and derives from an Anglo-Saxon form *gebed* with the same meaning. The semantic transfer in this form was effected by the general medieval European cultural practice of counting one’s prayers on rosary beads (still practiced within the Catholic church). Where $A =$ prayer and $B =$ bead, then, the semantic transfer in modern English can be schematised as $A \rightarrow A + B \rightarrow B$ (prayer $> \text{prayer + rosary bead} > \text{bead}$). Similar explanations underlie semantic transfers in the other two cases. English *pen* is a borrowing of Latin *penna* ‘feather’. After the Christianisation of Europe and the introduction of literacy writing began with the use of feather quills, an instrument of writing that was in use throughout Europe and the Americas until well into the 19th century. Then the advent of ink pens and later ball points put the feather quill to rest in the dustbin of history. The etymology of *pen*, however, tells us even without historical documentation, that feather quills were once used as writing implements. In categorial terms the meaning of *pen* went from $A =$ feather quill, to $A + B =$ feather quill + writing implement, to $B =$ writing implement (of any material). The example of *clock* is equally revealing of cultural changes in the history of Europe. English *clock* is cognate with German *Glocke* ‘bell; clock’, and Dutch *klok* ‘bell (of church, ship); clock’. Moreover, it derives from Anglo-Saxon *clugge* ‘bell’. Why a word that originally meant only ‘bell’ would come to mean ‘clock’ in modern English becomes clear when we recall the history of public time-keeping in Europe, in which the church bell was rung at each (approximate) hour of the day and night. Mechanical time-keeping devices were invented in Germany around 1500, and as these became more practical and available to the average person the reliance on public time-keeping by the ringing of church bells gradually passed out of use in most regions (although it still persists in parts of Europe). Again, we have a semantic change that can be characterised in categorial terms by the formulaic $A \rightarrow A + B \rightarrow B$, since the transition was from $A =$ bell, to $A + B =$ bell + time-keeper, to $C =$ time-keeper (of any type).

From the standpoint of the research questions central to cognitive linguistics it would be easy to say ‘These are instances of metonymy’ and leave the matter at that. But this would be missing the point that I am making with such examples. By all indications metonomy is at least as all-pervasive in daily language use as metaphor. Koch (1999:139) goes so far as to say ‘I am convinced that metonymy occurs much more frequently than metaphor and tells us a great deal about our cognitive equipment.’ Most examples of metonymy,

---

4 The case of *clock* raises further questions, as ‘Where does English *bell* come from?’. Again, there are cognates in Dutch *bell-en* ‘to ring’ and German *bell-en* ‘to bark, yelp’. The English noun *bell* thus apparently derives from a verb, and the sequence of relationships (*bell* $>$ *clock*, to *ring* $>$ *bell*) can be viewed as a kind of semantic ‘drag chain’.
however, tell us nothing about culture history, whereas those that can be represented by the formula $A \rightarrow A+B \rightarrow B$ often do.

What does any of this have to do with Austronesian languages? The examples of *bead, pen* and *clock* are valuable tests of Bloomfield’s formulaic $A \rightarrow A+B \rightarrow B$, since the inferences from semantic change can be tested against recorded history, but what about cases where we have no historical record to use as a metaphorical safety net? Here the value of inferences about semantic change must be more open to question, but their basic soundness follows from the structural integrity of the formula $A \rightarrow A+B \rightarrow B$, just as the basic soundness of the Comparative Method of linguistics follows from recognition that when applied to the Romance languages or other cases where documentary checks are available, it is found to yield results that are essentially correct. A rather undramatic example of semantic change in Austronesian languages that adheres to the same structural principles as English *bead*, *pen* and *clock* is seen in reflexes of Proto Malayo-Polynesian *tuzuq ‘to point, indicate’. Extremely widespread cognates make it clear that the PAn word for ‘seven’ was *pitu. However, in a large swath of Borneo and adjacent areas, including the Malay peninsula, the Chamic region of mainland Southeast Asia, eastern Sumatra, and western Java, *pitu has been replaced in the meaning ‘seven’ by reflexes of *tuzuq, as in Ida’an *turu?, Kadazan *tuu?, Kelabit *tudu?, Bintulu *tuju?, Kayan *tusu, Lahanan *tuju, Quop Land Dayak *ju?, Malay *tujoh, Rejang *tujua?, Jarai *tjuh, Tsat *su? (where the numeral superscript indicates a high-level tone), or Sundanese *tujuh. Why would such a semantic change take place?

Wilkinson (1959:1242) notes of Malay *tujoh ‘seven’ that it is ‘etymologically associated with the index finger … After counting all the fingers of the hand (*lima*), we come back to the thumb (six) and the index-finger (seven).’ In other words, the traditional method of serial counting was on the fingers of one hand, using the thumb as a pointer. By this method of counting the index becomes the seventh finger. If Wilkinson’s explanation is correct this semantic change reflects a cultural practice that probably is now waning, at least in urban areas, a practice that can be signaled by the ‘B’ element in the categorial sequence $A ‘index finger’ \rightarrow A+B ‘index finger + seven’ \rightarrow B ‘seven’.

A much more dramatic semantic change that follows the same structural principle is seen in reflexes of Proto Malayo-Polynesian *liaŋ ‘cave’, a meaning that can confidently be reconstructed on the basis of reflexes extending from the northern Philippines (Isneg *liyāŋ ‘cave, cavern’, Bontok *liyāŋ ‘cave’) to the Banks Islands (Mota *lia ‘hollow in or under a rock; cave, den’). In most languages spoken in northern Sarawak, however, reflexes of *liaŋ refer instead to places of burial, as seen in Long Anap Kenyah *liaŋ ‘grave, Long Wat Kenyah *liaŋ ‘cemetery, burial place’, Baram Kayan *liaŋ ‘burial post or grave’, Kelabit *liaŋ tanom ‘grave’, Batu Belah Berawan *liaŋ ‘wooden house-shaped coffin raised on pillars’ (with regular fortition of a predictable intervocalic glide and centralisation of the high vowel that triggered the glide), Long Teru Berawan *liaŋ ‘single-use post or pillar tomb’. It should be noted that in most cases where adequate descriptions are available the type of burial place described by a reflex of *liaŋ is an artificially elevated ossuary reserved for secondary interment of the remains of members of the upper class. Metcalf (1976), who has described these types of structures among the Berawan in considerable

---

5 Abbreviations of protolanguage names used in this paper are: PAn – Proto Austronesian; PMP – Proto Malayo-Polynesian; PPn – Proto Polynesian.

6 In reality, the numerals 4–7 would be indicated in the same way, with 4 = thumb on index, 5 and 6 = index on thumb, and 7 = thumb on index. The index finger is now called *tolunjuk* in Malay.
detail, distinguishes between the *salɔŋ* (a house-like structure raised on two ironwood pillars, with doors that permit reopening for further depositing of bones), and the *lijeŋ*, a niche in a single towering ironwood pillar that is sealed once and for all following secondary deposit of the bones of the deceased. The gloss that I recorded for Batu Belah *laŋ* suggests that the referent designated by reflexes of *liaŋ* may be either of these types of burial structure among different Berawan groups. Similar structures (called *kliŋ*) were described for the Sekapan and Kejaman of the upper Rejang river in the late 19th century by Roth (1896), where the massive ironwood pillars were carved with multiple niches to incorporate the bodies of slaves entombed to serve their chief in the afterlife. There is little physical similarity between such Bornean mausoleums and natural caves, and the fact that both referents are described with a reflex of *liaŋ* might be treated as coincidental.

A careful consideration of the entire linguistic comparison, however, takes us back to Bloomfield’s categorial progression. Throughout much of Sulawesi cave burial was practiced within the ethnographic present, as seen in the gloss of Tae’ (Southern Toraja) *liaŋ* ‘cave burial in the face of a cliff’. Similarly, as noted in Blust (1986/87), cave burial was described in the late 19th century for many of the peoples of eastern Sabah, and early in the 20th century for the Penihiing (Kenyah) of the Mahakam river in Central Kalimantan, the Busang (Kayan) of the upper Kapuas river in West Kalimantan, and others. What is striking is that cave burial was not practiced within the ethnographic present by any of the peoples who use the elaborate mortuary structures described above. Nonetheless, a sequence of radiocarbon dates from the Great Cave at Niah in northern Sarawak, shows that extended burials in coffins or wrapped in matting were made in this natural limestone cavern until around 1000 AD (Brooks et al. 1979:30). Given this larger context it seems clear that reflexes of *liaŋ* meaning ‘cave’ and reflexes of *liaŋ* meaning ‘mausoleum on an ironwood post’ are historical continuations of one and the same term, and that the meaning of *liaŋ* and its various reflexes in northern Sarawak passed through a categorial progression $A \rightarrow A+B \rightarrow B$, where $A$ = cave, $A+B = cave + burial place$, and $B$ = burial place of any type. Superficially this change differs sharply from the semantic history of English words such as *bead*, *pen* or *clock*, but in terms of categorial linkage and decoupling all four exemplify the same type of semantic evolution. Cases of this type, then are examples not just of metonymy as a general classificatory label, but of a special subtype of metonymy that allows historical linguists to extract important information about culture history from the evidence of semantic change.

2.2 Semantic fragmentation

One of Bloomfield’s nine types of semantic change is described (1933:427) as ‘synecdoche — the meanings are related as whole and part’, and is illustrated with ‘Primitive Germanic *[“tu:naz* ‘fence’ (so still German *Zaun*) > *town*’.

Since synecdoche is normally understood to be a device in which a part represents the whole this example again seems ill-chosen, since if the original sense was ‘fence’ the meaning of English ‘town’ would show an expansion of reference to a more inclusive meaning. I believe there is a better explanation for the semantic development seen in the English word ‘town’, and that this explanation involves a mechanism that is common to semantic changes in many Austronesian languages. First, it must be recognised that this cognate set also includes Dutch *tuin* ‘garden’, and Irish *dun* ‘fortification’. How can we make sense of this range of meanings (town, fence, garden, fortification) in association with a single etymon?
Many cities of Europe and Asia still preserve the old town walls that once surrounded them in times when security was achieved by shutting out potential attackers. Early and medieval towns in Europe, then, were surrounded by protective walls, and the idea of a town was inseparable from the idea of a walled-in enclosure. This type of referent can be thought of as a ‘semantically complex’ notion, that is, as a notion which can be decomposed into constituent elements with a high degree of independence (towns can exist without walls, and walled enclosures need not be towns). From this complex of semantic components German separated out the notion of an enclosing structure, English separated out the notion of the resident population and buildings within the enclosing structure, Dutch transferred the idea of an enclosing structure to the domain of plants, and Irish preserved the original sense with only minor change. Examples such as *town can be thought of as illustrating ‘semantic fragmentation’ — components of a complex meaning are separated and transmitted individually into different daughter languages. An example of this type of semantic change in Austronesian languages is seen in Table 1.

| **Taiwan** | | | |
| --- | --- | --- |
| Kavalan | balat | east wind |
| Amis | safalat | south wind |

| **Philippines** | | | |
| --- | --- | --- |
| Itbayaten | havayat | west wind |
| Ilokano | abágat | south wind |
| Casiguran Dumagat | abágat | South-west monsoon winds |
| Tagalog | habágat | west or south-west wind; monsoon |
| Western Bukidnon Manobo | evagat | North-east monsoon |
| Tiruray | barat | rainy season |

| **Western Indonesia** | | | |
| --- | --- | --- |
| Malay | barat | west |
| Malagasy | avaratra | north |
| Old Javanese | barat | strong wind, storm; west |
| Sasak | barat | storm; to storm |
| Tae’ | baraʔ | big, terrific, violent, of rain and wind |

| **Eastern Indonesia** | | | |
| --- | --- | --- |
| Manggarai | warat | rainy season |
| Roti | fa(k) | west monsoon, rainy season |
| Savu | wa | west; the island of Sumba |
| Buli | pāt | west, west wind |
| Numfor | barək | west |
Five patterns of semantic change in Austronesian languages

<table>
<thead>
<tr>
<th>Oceania</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wuvulu</td>
<td>afaa</td>
<td>northwest wind</td>
</tr>
<tr>
<td>Likum</td>
<td>yahay</td>
<td>west wind</td>
</tr>
<tr>
<td>Mussau</td>
<td>apace</td>
<td>strong wind, storm wind</td>
</tr>
<tr>
<td>Manam</td>
<td>awara</td>
<td>northwest wind</td>
</tr>
<tr>
<td>Motu</td>
<td>lahara</td>
<td>northwest wind and season</td>
</tr>
<tr>
<td>Fijian</td>
<td>cavā</td>
<td>hurricane, windstorm</td>
</tr>
<tr>
<td>Samoan</td>
<td>afā</td>
<td>storm, gale, hurricane</td>
</tr>
<tr>
<td>Maori</td>
<td>awhā</td>
<td>gale, storm, rain</td>
</tr>
</tbody>
</table>

Together these widespread cognate forms make it clear that in Proto Malayo-Polynesian *habaRat referred to the west or north-west monsoon (the matter may have been different in PAn, as this community would have been just beyond the normal reach of the monsoons, and even in Malayo-Polynesian languages the directionality clearly differs with latitude). However, if the comparison were limited to, say, Tiruray barat ‘rainy season’, Malay barat ‘west’, Fijian cavā ‘hurricane, windstorm’ the meaning of this term would be far more difficult to determine. This point can be made even more forcefully with the complementary term PAn *timuR ‘south wind’(?), PMP *timuR ‘south-east monsoon’ which shows such diverse reflexes as Kavalan timuR ‘south wind’, Amis timol ‘south’, Tagalog tīmog ‘south, south wind’, Timugon Murut timug ‘water’, Malay timur, Komodo tīmu ‘east’, Roti tīmu ‘east, east monsoon’, Buli simi ‘south, south wind’, Samoan tīmu ‘be rainy’, East Futunan tīmu ‘buffeted by wind’, Anuta tīmu ‘rain lightly, drizzle’, and Rennellese tīmu ‘to rile, devastate, as by wind and storm’. The meaning of both PMP *habaRat and *timuR, then, can be seen as containing components of directionality, rainfall and wind, and these were differentially segregated in the histories of individual languages. In some cases this separation of semantic components happened in languages that continued to be spoken within the monsoon regime; in others (as the Polynesian languages), it can be argued that semantic fragmentation was triggered by migration outside the region in which the monsoons are known.7

It might be objected that meaning is inherently componential — there are no ‘simple’ meanings. Such an objection, however, reflects the limited vision of someone concerned solely with synchronic approaches to semantic structure. Although no work within the cognitive linguistics paradigm appears to recognise it, diachronic approaches to semantic change show unambiguously that some lexical categories are ‘complex’ in ways that others are not. Blust (1980, 1987), for example, has drawn attention to differences of ‘semantic profile’ as the key to a disciplined approach to semantic reconstruction in historical linguistics. To choose one of many possible illustrations, PAn *Rumaq meant ‘house, family dwelling’. Its reflexes in hundreds of languages are almost uniformly consistent in preserving this sense, extending it only metaphorically to the nests of birds, the webs of spiders, and the sheaths of knives. PMP *banua, on the other hand, had a meaning that is impossible to represent by a single English word. The *banua evidently was the territory that a community inhabited, including its sources of food and drinking water, and the graves of the ancestors (whose beneficence was essential to the wellbeing of the living).

7 Ross (1995) adds ‘island’ as a further semantic development in some of the languages of south-east New Guinea. This, however, is independent of the broader pattern of semantic change in which the monsoon terms underwent innovations that can be characterised as instances of semantic fragmentation.
It was, in effect, the territorial unit that formed the life-support system of a human community. As a result, semantic reflexes of this term are very diverse, including ‘landing place’, ‘year’, ‘sky/heaven’, ‘town/village’, ‘country/land’, ‘commoner/serf’, ‘thunder’, ‘sea’, ‘weather’, ‘house’, ‘night’, ‘field’, ‘mainland (as opposed to island)’, and ‘island’. To suggest that all meanings are ‘complex’ simply fails to capture the striking difference in the semantic evolution of partially overlapping terms such as PAN *Rumaq (which has a simple diachronic ‘semantic profile’) and PMP *banua (which has a complex diachronic ‘semantic profile’), although reflexes of both of these forms mean ‘house’ in widely separated Austronesian languages.

In the terms proposed by Seto (1999) most examples of semantic fragmentation would be classified as examples of metonymy, since they involve a ‘part of’ rather than a ‘kind of’ relationship of inclusion. English ‘town’, German Zaun ‘fence’, Dutch tuin ‘garden’, and Irish dun ‘fortification’ evolved from a word which referred to towns that implicitly were walled. In the first three languages only a part of this complex notion has been retained (and in Dutch transferred from human to botanical reference). In Irish the structural integrity of the original concept has been retained, but to the extent that fortifications came to be distinguished from towns a change of function is evident. With PMP *banua the richness of metonymic connections is even greater, extending to beliefs about the afterworld as a ‘shadow banua’ of the world of the living, as seen in Kapampangan banwa ‘year; sky, heaven’, Toba Batak banua ginjan ‘upper world, heaven’, banua toja ‘middle world, Earth’, banua toru ‘underworld, world of the dead’, Nias banua ‘sky, heaven’, or Mentawai manua ‘sky, heaven’. However, it is not clear that all examples of semantic fragmentation can be characterised as examples of metonymy.

English thatch is cognate with German Dach, Dutch dak ‘roof’. The semantic transfer again seems fairly straightforward: all roofs were thatched in early Germanic houses, and the component material has been separated out in English, but the organising structure in German and Dutch, in which ‘thatch’ must now be expressed by a compound: German Dachstroh, Dutch dakriet. Is this example of semantic change a case of metonymy or of synecdoche? In some ways this traditional method of classification seems inappropriate. In Seto’s terms the semantic change in English thatch might be considered a case of synecdoche, since a thatched roof is a kind of roof. But English ‘thatch’ refers to the material, whether it already forms a roof or not, and even when it does form a roof the thatch is not ‘part of’ the roof: it is the roof conceived as substance rather than form. The semantic change in German and Dutch is equally difficult to classify as a product of either metonymy or synecdoche, since a roof as a structure has neither a ‘part of’ nor a ‘kind of’ relationship to ‘thatched roof’ (where ‘thatched’ is implicit, since no other kinds of roofs existed at the time to allow a contrast of types). On balance, then, the notion of semantic fragmentation appears to account better for this set of data than traditional schemas of semantic change or more recent revisions of them within the cognitive linguistics paradigm.

Other examples of semantic fragmentation can be found in Austronesian languages as well, most or all of which affect monomorphemic terms that have no categorial equivalent in English. One of these is PMP *qayam ‘domesticated animal’, a term that can be

---

8 Unlike northern European societies, where Mediterranean influence introduced new types of roofing material in major population centers before the Middle Ages (and left traditional thatched roofs in the countryside), a similar process of semantic fragmentation has not affected PMP *qatep ‘(thatched) roof’, presumably because until very recent times all roofs in Austronesian-speaking societies were thatched.
reconstructed despite the absence of a viable etymon for the more general category ‘animal’ (Blust 2002). Semantically intact reflexes of this term (often in the sense ‘pet’) are found from the northern Philippines (Gaddang ayam ‘domesticated animal’), to the southern Philippines (Tiruray ayam ‘domesticated animal’, Sarangani Manobo ayam-ayam ‘pet, domesticated animal’), to Borneo (Kadazan tazam ‘tame animals’, Iban ayam ‘pet’), to Palau (Palauan charm ‘animal’, but ou-charm ‘keep an animal as a pet, raise an animal’). In a number of other languages cognate forms refer to particular domesticated animals, as with Tagalog, Bikol, Aklanon áyam ‘dog’, Pamona (Bare’e) ajá ‘substitute term for ‘dog’ (asu)’, Murik ayam ‘domesticated pig’ (cf. mabi ‘wild pig’), Malay, Old Javanese (h)ayam, Sundanese hayam ‘domestic fowl, chicken’ (can refer to wild fowl, but only with a qualifying word indicating ‘forest’ or the like). Less commonly, reflexes may be denominal verbs that derive from an earlier animal name, as with Cebuano áyam ‘for a dog to chase something; hunt using dogs’ (< earlier *ayam ‘dog’). Here the notion ‘domesticated animal’, which covered pigs, dogs and chickens in PMP society, has fragmented into individual members of the set in some languages, while retaining its integrity as a broader animal category in others.

Probably the richest set of examples of semantic fragmentation in Austronesian languages is found in the realm of kinship. This should come as no surprise, since kinship terminology was the semantic domain used to pioneer componential analysis in linguistic anthropology half a century ago (Goodenough 1956). Componential analysis chose its lexical domain on the basis of the ease with which it was possible to separate out components of meaning that are capable of independent definition, and of recombination in other semantically complex morphemes, and it is precisely this ease of separability that allows semantic fragmentation in the reflexes of many kin terms. One example is PAn *ma-tuaS, PMP *ma-tuqah, which at least by PMP times evidently meant ‘mother’s brother/wife’s father’ as a consequence of preferential matrilateral cousin marriage (Blust 1980, 1994). Reflexes of this term also have a variety of kin-related (and sometimes non-kin-related) meanings, including ‘parent-in-law’ (Thao tuq-tuqash, Mukah Melanau matua, Ba’amang mantuha?, Malay montualmortua, Toba Batak si-matua, Manggarai, Central dialect tuʔa, ‘parents of the bride’ (Nias matua, Manggarai ‘SH’ dialects tuʔa), ‘parent’ (Proto Polynesian *matua), ‘old, adult, full-grown’ (Kambera matua, standard Manggarai tuʔa, Rotuman mafua), ‘big, large’ (Roti matua), ‘leader, spokesperson’ (Malay ke-tua, Rotuman mafua), ‘traditional class of nobles’ (Chamorro matua), and ‘mother’s brother’ (POC *matuqua). The semantic components in the inferred meaning of PMP *ma-tuqah include +1 generation, +collateral, +male, +consanguineal, +affinal, +important (‘big, large’, ‘leader, spokesperson’, ‘class of nobles’), and again all of these have been segregated in the semantic reflexes of particular forms. Other examples could be given, but these should be sufficient to make the point that semantic fragmentation is a fairly common type of change that is sometimes difficult to characterise as metonymy.

2.3 Semantic change due to change of environment

When a language community enters a new environment, it faces the challenge of finding lexical labels for unfamiliar referents, particularly those relating to the natural world. These can be coined de novo, but economy of effort favours the novel use of inherited forms. Most Austronesian speakers are exposed to a similar climate between the Tropic of Cancer and the Tropic of Capricorn, since the major thrust of the migrations out of Taiwan into insular Southeast Asia and out into the Pacific was from west to east.
Probably the most notable exception to this pattern is that of the Maori, where a community that had developed for centuries in a tropical climate settled lands as far south as 47 degrees south latitude, and so found itself surrounded by many unfamiliar flora and fauna. Clark (1982) has shown how some Proto Polynesian bird names were redefined in the New Zealand context by shifting from one avian referent to another, as with PPn *kiwi ‘a shore bird, possibly the curlew or sandpiper’ > Maori kiwi ‘flightless bird of the genus Apteryx (similar to curlews and sandpipers in extracting worms from sandy soil with a long beak)’, and most strikingly, with PPn *moa ‘chicken, domestic fowl’, a term that came to apply to the eleven species of the genus Dinornis, a group of endemic flightless birds the largest of which stood ten feet tall and weighed in at 500 pounds. Here the basis for the shift of referent is unclear, although both the weakly developed flying ability of domestic fowls and the ease with which the naïve native moa was hunted for food probably paved the way. More striking still, because it involved a transfer of reference from a marine reptile to a bird, is PPn *kea ‘hawksbill turtle’, a term that came to be applied instead to a large carnivorous dull-green parrot endemic to New Zealand. Here both the colouration and the beak-like mouth of the turtle undoubtedly played a part in determining the use of this word for the newly encountered bird.

Another parameter of environmental change is altitude, since change of altitude is often homologous with change of latitude. In some of the languages of the Central Cordillera of northern Luzon reflexes of PAn *daya ‘upstream, toward the interior’ have come to refer to the sky either physically, or as a mythological region of the gods, as in Bontok dáya ‘sky; heaven’, Kankanaey dáya ‘heavens, firmament, sky, used only in tales’, Ifugaw dáya ‘cosmic upstream region extending beyond the sky cupola … the cosmic upstream region is inhabited by the gods and spirits of the Dáya’, Batad Ifugaw dáya ‘the heavens (the area includes the earth’s atmosphere, touched by mountain peaks, where birds fly; it is the location of clouds, and the source of meteorological events, as rain, hail, thunder and lightning, storms and the like. The territory also includes outer space and involves spirits relating to, and named after various celestial bodies, such as the sun, moon, numerous named stars, sunlight, and so forth.’ What is noteworthy about this change is that it is known only in languages spoken by populations that are located at the highest elevations in the Cordillera: having gone upriver as far as they could, the inherited term that once meant ‘upriver’ could only be projected onto the firmament.

Other areas that differ not so much in latitude or altitude, as in a geological history of isolation that produced many endemic species of plants and animals, include New Caledonia in the Pacific, and Madagascar in the western Indian Ocean. For various reasons (including their low cognate densities with languages elsewhere) little can yet be said about semantic change in the languages of New Caledonia. Although Malagasy is relatively well-studied and has fairly high cognate densities with other Austronesian languages, virtually no examples of semantic changes triggered by exposure to a novel physical environment such as that in New Zealand can be cited (one might have expected, for example, that inherited terms for various monkeys, which are absent in Madagascar, would have been applied to the newly-encountered lemurs). The contrast between the relatively frequent occurrence of semantic change in relation to the faunal environment in

---

9 In the case of New Zealand it was not just the change of latitude that confronted the Maori with new flora and fauna, but the longstanding geological isolation of these islands from other major landmasses, giving rise to many unique species.
New Zealand and the rarity of semantic change in relation to the faunal environment in Madagascar is thus surprising.

Perhaps the most striking semantic change motivated by change of physical environment known in Austronesian languages is that of the PMP doublets *taRutuŋ and *taRutum ‘porcupine fish’, and here the crucial factor inducing semantic change is not change of latitude or altitude as such, but rather movement inland. Evidence for the term with this meaning is given in Table 2.

Table 2: Evidence for PMP *taRutuŋ and *taRutum ‘porcupine fish’

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Meaning</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cebuano</td>
<td>tagutúŋ-an</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Sasak</td>
<td>kərutuŋ (*t &gt; k irreg.)</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Palauan</td>
<td>darúdm</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Waay (Ambon)</td>
<td>tarutum</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Hila (Ambon)</td>
<td>tarutun</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Kiribati</td>
<td>tauti</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Rotuman</td>
<td>faufu</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
<tr>
<td>Proto Polynesian</td>
<td>*tautu</td>
<td>porcupine fish</td>
<td><em>Diodon</em> sp.</td>
</tr>
</tbody>
</table>

Since reflexes of these terms are distributed from the central Philippines to Polynesia in the meaning ‘porcupine fish’, it must be assumed that this meaning is original. In a number of daughter languages, however, reflexes refer instead to the porcupine, as in Lun Dayeh, Bario Kelabit tarutuŋŋ, Katingan tahatuŋŋ, Maanyan tetuŋŋ, Komodo, Manggarai rutuŋŋ, Sanggar (Bimanese dialect) tarutuŋ ‘porcupine: *Hystrix javanica*’. Even more striking, in other languages reflexes refer to thorny plants, as with Karo Batak, Toba Batak tarutuŋ ‘durian’, or Pohnpeian sei ‘porcupine fish; soursop: *Annona muricata*’. It is noteworthy that all reflexes of *taRutuŋ in which the meaning ‘porcupine fish’ has been completely lost are found in languages that are located either in the interior of large islands (Lun Dayeh, Bario Kelabit, Katingan, Maanyan, Karo Batak, Toba Batak) or in the Lesser Sundas, where the traditional cultures have an aversion to the sea and are land-oriented even on small islands (Komodo, Manggarai, Sanggar). These recurrent semantic changes appear to have been triggered, then, by a physical or cultural movement away from the sea, leading to loss of direct contact with the original referent, and application of the retained term either to a land mammal or to fruits that bear a striking physical resemblance to the inflated body of the porcupine fish.10

---

10 As the Pohnpeian example shows, extension of reference to thorny fruits or plants did not require the prior loss of reference to the porcupine fish. Madulid (2001:690) lists additional plant terms such as ‘Bisayan’ (not further specified) taguṭoŋ, Mangyan (Batangan) taguṭuŋ, ‘Solanum ferox L.’, Mangyan taguṭoŋ ‘Solanum torvum Sev.’, Bikol taguṭoŋ-an ‘Euphorbiaceae: *Bridelia penangiana* Hook.’. The *Solanum ferox* L., also known as the hairy eggplant, reportedly has spines capable of producing injury, and the *Solanum rostratum* is commonly known as the buffalo burr, since its spiky seedpod often sticks in the fur of passing quadrupeds.
2.4 Prototype/category interchange

In some ways it can be claimed that most or all typologies of semantic change since Bréal are Eurocentric, since nearly all examples that they cite are from the Indo-European languages of Europe, and the categories they propose seem to offer both too much and too little to accommodate the data of Austronesian languages. Moreover, some semantic innovations in Austronesian languages suggest that traditional typological categories such as ‘widening’ and ‘narrowing’ of meaning may be instances of the same type of change.

To illustrate, PMP had the words *hulaR ‘snake (generic)’ and *sawa ‘python’, but in a number of daughter languages reflexes of *sawa mean ‘snake (generic)’: Yakan sawa, Dampelas saa, Balaesang ule saa, Ratahan, Boano, Simalur, Bimanese sawa ‘snake (generic)’. Innovations for ‘python’ include Yakan bəkəssan, from *beokes ‘wrap tightly or firmly by binding’, Dampelas saa babate, and Boano sawa marimba, the latter two of obscure origin (Himmelmann 2001:150). The semantic change ‘python’ > ‘snake’ in connection with reflexes of *sawa could be forced into Bloomfield’s typology as an example of widening of meaning. To do this without further discussion, however, would obscure the fact that these examples represent at least five historically independent semantic innovations (one in Yakan, another in Ratahan, a third in Dampelas, Balaesang and possibly Boano, a fourth in Simalur, and a fifth in Bimanese). Recurrent semantic changes have received no attention in the Eurocentric literature from Bréal through the most recent work of scholars within the cognitive linguistics paradigm, yet they provide crucial evidence that some aspect of perception or psychology has been active through generations of speakers, favoring one type of change over others that might have occurred, but didn’t. The reticulated python is the largest and psychologically the most impressive snake in insular Southeast Asia, and the equation python = snake suggests that it was conceived as the prototype of the snake category, the snake par excellence. In this respect it resembles the equation ‘cottonwood’ = ‘deciduous tree’ in New Mexican Spanish and a number of the native languages of the American Southwest (Trager 1939), or the equation ‘eagle’ = ‘bird’ in a number of Uto-Aztecan languages, including at least Hopi and Proto Numic (Hage and Miller 1976).

Although well camouflaged by a complex historical phonology, Palauan ȵūis ‘green tree viper, the Palau tree snake, *Dendrelaphis lineolatus*’ reflects *hulaR*, and superficially appears to involve a process of semantic transfer that is the reverse of that in reflexes of *sawa* (from general to particular). Most sources state that there are five snakes in Palau. Two of these are venomous but non-aggressive, with no history of snakebite casualties. A third is reclusive and rarely seen, and the fourth is the Brahminy blind snake, a diminutive species that spends much of its time underground. Since the python is absent in Palau it is a reasonable inference that the green tree viper, which is described as a ‘very fast, nervous snake’ common in small trees and shrubs, has usurped its place as the most dangerous and hence psychologically most salient snake in these islands. In Bloomfield’s terms this semantic change could be classified as an example of narrowing of meaning, but given the larger context sketched here both this and the preceding change can be seen as sharing a common basis. Whether a change is from prototype to category (as with *sawa*), or from category to prototype (as with *hulaR*), the category/prototype boundary becomes blurred and eventually lost. In one case this appears to be an instance of semantic widening, and in the other an instance of narrowing, but in both cases a semantic change arguably occurs because a prototype comes to be identified with the entire category of which it is a part.
Given these two examples that involve snakes it is worth asking whether a similar fusion of category and prototype is known in the semantic history of other faunal terms. The answer is tentatively negative. Whereas the monitor lizard (PMP *bayawak) is by far the largest and psychologically most prominent lizard in insular Southeast Asia and the western Pacific, it differs from the python in having no real competitors (all other lizards in this part of the world are so miniscule by comparison that Varanus species in many ways stand apart as a different type of animal than the geckos, grass lizards and the like that constitute the remaining lizard population). Pythons, then, are seen as the monarchs of the snake kingdom, but monitor lizards probably are not compared with other lizards. For birds, mammals, fish, or any part of the plant kingdom there is no clearly dominant member that could trigger a prototype-category fusion of the sort seen with reflexes of *sawa and *hulaR.

### 2.5 Avoidance

The last category I will discuss is avoidance. Indo-Europeanists are familiar with the various circumlocutions that have been applied to the bear out of fear and respect (Buck 1949). While this may be considered lexical innovation rather than semantic change, the motivation is roughly the same. When speakers of Iban substitute the word ‘vine’ for ‘snake’ while traveling through the jungle (but not otherwise), it is to minimise the chance that uttering the name will attract the creature. While this has not yet led to a semantic change ‘vine’ > ‘snake’ in this language, the appearance of both meanings with a reflex of PMP *waRej ‘vine, creeper’ in Maranao wagød ‘snake; vine’, and Buruese wahst ‘vine; snake’ is almost certainly due to the same use of substitute terms in situations of perceived danger, and hence is motivated by a desire to avoid attracting or confronting a dangerous animal. Slightly different is the practice, in some parts of Indonesia, of addressing a crocodile as ‘grandfather’, evidently as an act of appeasement to forestall a potential attack.

One semantic category for which lexical innovations of a particular type are recurrent is the word for ‘blood’. While PAn *daRaq ‘blood’ can be reconstructed with complete confidence, many daughter languages have replaced this term with a word that originally meant ‘sap, juice, or the like’, as shown in Table 3.

#### Table 3: Replacement terms for PAn *daRaq ‘blood’

<table>
<thead>
<tr>
<th>Language</th>
<th>Replacement</th>
<th>Original Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagalog</td>
<td>dugó?</td>
<td>blood</td>
</tr>
<tr>
<td>Hanunóo</td>
<td>dugú?</td>
<td>blood</td>
</tr>
<tr>
<td>Maranao</td>
<td>rogo?</td>
<td>blood</td>
</tr>
<tr>
<td>Subanun</td>
<td>dugu?</td>
<td>blood</td>
</tr>
<tr>
<td>Bolaang Mongondow</td>
<td>dugu?</td>
<td>blood</td>
</tr>
<tr>
<td>Gorontalo</td>
<td>duhu</td>
<td>blood</td>
</tr>
</tbody>
</table>

11 The largest lizard in insular Southeast Asia apart from species of the genus Varanus evidently is the nocturnal tokay gecko of the Greater Sunda islands, which reaches a length of about fourteen inches but is shy and retiring, its presence often known only from its distinctive clicking chirp.
(2) PMP *pulut ‘breadfruit sap; gluey substance’

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sebop</td>
<td>pulut</td>
<td>sap of a tree; blood</td>
</tr>
<tr>
<td>Kenyah (Galvin 1967)</td>
<td>pulut</td>
<td>rubber tree; sap, sticky; euphemism for blood</td>
</tr>
</tbody>
</table>

(3) PMP *liteq ‘sap of a tree’

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilaan</td>
<td>litoʔ</td>
<td>blood</td>
</tr>
</tbody>
</table>

(4) Proto Greater Central Philippines *tag ‘sap’

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamian Tagbanwa</td>
<td>tagɐk</td>
<td>blood</td>
</tr>
<tr>
<td>Palawan Batak</td>
<td>tagɐk</td>
<td>blood</td>
</tr>
</tbody>
</table>

(5) Lou

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>turI</td>
<td>blood; sap of a tree</td>
</tr>
</tbody>
</table>

(6) PMP *giteq ‘sap of a tree’

<table>
<thead>
<tr>
<th>Language</th>
<th>Form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Javanese</td>
<td>gətih (met.)</td>
<td>blood12</td>
</tr>
<tr>
<td>Balinese</td>
<td>gətih (met.)</td>
<td>blood</td>
</tr>
</tbody>
</table>

Other possible examples of the same type of change are Giman (southern Halmahera) git ‘blood’ (< *giteq?), and Dusner (West New Guinea) rikat ‘blood’, which may reflect PMP *diket or *riket ‘sticky, adhesive’. Although it is relatively stable, Pan *daRaq has, of course, been replaced by many other lexical items, most of which are of unknown origin. In a few cases a semantic change path can be determined, as with Northern Kankanaey basá ‘blood’ (< *baseq ‘wet’), or the Greater Central Philippines form *laŋesa ‘odour of fish or blood’, which became the ordinary word for ‘blood’ in Proto Manobo. However, cases like this are isolated, and do not form the kind of recurrent pattern seen with ‘tree sap’ > ‘blood’ repeatedly in the history of Austronesian languages.

At first the pattern of semantic innovation seen here may seem like the types of cases considered by Brown and Witkowski (1981b): certain change paths are universally favored in the formation of lexical categories, and so will be followed repeatedly in different languages and language families. There is, however, possibly a difference in this case, since in some languages, as Sebop and Lou, a single morpheme means both ‘sap of a tree’ and ‘blood’, and for at least the Kenyah languages of Sarawak this form is sometimes said to be a euphemism for ‘blood’. Euphemisms are essentially devices for avoiding direct reference to something that is considered embarrassing, unpleasant or dangerous, and the use of this mechanism in Kenyah pulut ‘sap, sticky; euphemism for blood’ raises the

12 Modern Javanese gətih occurs in Ngoko, the common speech register (cf. rah < *daRaq in Krama, the high register). Old Javanese, representing an east Javanese literary tradition preserved in palm leaf manuscripts dating from the 9th to 15th centuries, has gətih ‘sap’, showing that this semantic innovation has taken place within the past millennium. The Balinese form is almost certainly a loan from Javanese.
possibility that many of these examples began as alternative means of expression for a morpheme that was avoided as a result of its association with danger, slaughter or death.\footnote{Blood has a purifying virtue in many Southeast Asian cultures, where animal sacrifice is accompanied by the sprinkling of the victim’s blood in an act of spiritual cleansing or atonement. This, however, is animal blood, and is best seen as part of an offering to the unseen powers that control human fate. By contrast, human blood is more commonly regarded as a sign of danger, an interpretation that apparently is needed to explain why this word has so often been replaced by morphemes representing the same category of meaning — a phenomenon that is not known to be paralleled by other words for body parts or fluids.}

### 2.6 Conclusions

For well over a century the study of semantic change has been based almost entirely on data from Indo-European languages. The initial interest was in constructing typologies which would allow all semantic changes to be categorised into a limited number of types. More recent work in cognitive linguistics has concentrated more intensively on limiting the categories (only metaphor, metonymy and synecdoche are generally recognised), and on clearly distinguishing them from one another. Data from both Indo-European and Austronesian languages sometimes can be accommodated comfortably within these schemas, but doing so fails to bring out special qualities that they share, as with semantic changes that ‘freeze’ certain aspects of culture history in the semantic history of words (English bead, pen, bell, reflexes of Proto Malayo-Polynesian *liaŋ ‘cave’ in a number of the languages of northern Sarawak). In other cases, as with ‘semantic fragmentation’, there appears to be no natural way to express these changes as instances of metonymy or synecdoche, and they therefore suggest that current schemes of classification are too narrow and rigid to accommodate all known examples of semantic change in natural languages.

### References


Samoan was the first Polynesian language I knew anything about. I learned its basic vocabulary as I would that of any completely unfamiliar language. Before long, however, my impulse toward comparative linguistics led me to peek at other Polynesian languages, and gradually I began to realise that they were all out of step except Samoan! I had learned the Samoan word *maile* for ‘dog’, but in Tongan ‘dog’ was *kulii*, in Maori *kuri*, and so on. Words of the form *maile* elsewhere referred to a sweet-scented plant. A couple of other items emerged when I recorded a traditional story about a shark and a turtle. In Samoan they were the *malie* and the *laumei*; but in almost any other Polynesian language these sea creatures would be referred to as, respectively, *manoo* and *fonu*. The Samoan shark-word seemed to mean something like ‘nice’ in other languages, and the turtle-word appeared to refer to some kind of leaf.

In time I came to accept that the non-Samoan forms continued the original vocabulary of Proto Polynesian (PPn). The phenomenon of ‘Samoan specialties’ in the lexicon continued to intrigue me, but it was many years before I took a more systematic look at it. In the present paper I present a sample of these words, and one or two suggestions of factors that may have been at work in their formation.

This anecdotal introduction is to some degree necessary, since it is difficult to prove rigorously that there is anything unusual going on in Samoan. It could be that my perception of these distinctive words is distorted by the way in which I first encountered them. Lexical replacement is, after all, part of normal linguistic change. And when seen within its broader Austronesian context, Samoan is lexically quite a conservative language. In Blust’s study of the rate of retention of Proto Malayo-Polynesian basic vocabulary in a large sample of daughter languages (Blust 1981), Samoan, with 42 per cent, was the second most retentive Oceanic language, and outranked all the other Polynesian languages in the sample (Tongan, Futuna, Mele-Fila, Maori, Hawaiian, Tahitian) (Robert Blust, pers. comm.). Polynesian languages are, as a group, lexically retentive as Oceanic languages go.

1 By the time I learned my first few words, Andy Pawley was already a fluent speaker of Samoan and had published a number of papers on it, including his MA thesis (Pawley 1966). In presenting this small tribute to a long-time colleague and friend, it is a pleasure to draw attention to this less-known early aspect of his work.

2 The first version of this paper was presented at the First International Conference on Oceanic Linguistics in Port Vila, Vanuatu, 1993.
Even within Polynesian, Samoan does not appear highly innovative in aggregate. In an informal study, I compared half a dozen Polynesian languages for retention of Proto Polynesian vocabulary, using the same basic vocabulary list as in Blust’s study. Only Tuvaluan showed fewer replacement items than Samoan. Many of the ‘Samoan specialties’, however, occur nowhere else, not being shared even with close relatives such as Tokelau and Tuvaluan. It would appear, then, that after a long history characterised by slow replacement, the very recent history of Samoan has seen a number of unique replacements, affecting even such highly stable lexical items as ‘die/dead’, ‘left’, ‘long, rain’ and ‘sea’. This of course is not to rule out the possibility that a careful examination of the lexicon of other languages would reveal similar patterns of replacement.

Some Samoan lexical replacements are illustrated below in six groups, according to the form and apparent origin of the replacement word.

(1) In the first group, a PPN source for the replacement is fairly easy to see, but a significant semantic shift has taken place.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPN</th>
<th>Samoan</th>
<th>Source of Samoan</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘armpit’</td>
<td>*afiga</td>
<td>’ao’ao</td>
<td>*kahokahoe ‘side, rib’</td>
</tr>
<tr>
<td>‘banana’</td>
<td>*futi</td>
<td>fa’i</td>
<td>*faki ‘pluck, break off’</td>
</tr>
<tr>
<td>‘candlenut (Aleurites)’</td>
<td>*tuitui</td>
<td>lama</td>
<td>*rama ‘torch’</td>
</tr>
<tr>
<td>‘cooked’</td>
<td>*maoha, *moso</td>
<td>vela</td>
<td>*wela ‘burning’</td>
</tr>
<tr>
<td>‘dead (animal)’</td>
<td>*mate</td>
<td>pē</td>
<td>*peqe ‘over-ripe, rotten’</td>
</tr>
<tr>
<td>‘dead (person)’</td>
<td>*mate</td>
<td>oti</td>
<td>*qoti ‘finished’</td>
</tr>
<tr>
<td>‘fall’</td>
<td>*too, *siga</td>
<td>pa’ū</td>
<td>*pakuu ‘go boom’</td>
</tr>
<tr>
<td>‘fear’</td>
<td>*mataku</td>
<td>fefe</td>
<td>*fefe ‘shrivelled’</td>
</tr>
<tr>
<td>‘long’</td>
<td>*loa</td>
<td>’umi</td>
<td>*kumi ‘ten fathoms’</td>
</tr>
<tr>
<td>‘rain’</td>
<td>*quha</td>
<td>timu</td>
<td>*timu ‘squall’</td>
</tr>
<tr>
<td>‘skin’</td>
<td>*kili</td>
<td>pa’u</td>
<td>*paku ‘scab, crust’</td>
</tr>
</tbody>
</table>

(2) In the second set, while the Samoan words are unattested elsewhere with the same or similar meaning, there do exist homophones (or near-homophones) from which they could conceivably be derived. In some cases the semantic relations are, to say the least, obscure, and the resemblances may be fortuitous. However, one might imagine a connection between ‘full’ and ‘top’; and perhaps some sort of apotropaic magic at work in applying a term like malie/mālie to the shark.

---

3 The other languages compared were Tongan, Rennellese, Hawaiian and Maori.
4 Samoan words cited throughout follow the standard orthography in using <’> for the glottal stop, <g> for the velar nasal, and a macron <’> To mark phonemically long vowels. In reconstructed forms <g> has the same value but <q> is used for the glottal stop.
Table 2: Homophonic items

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
<th>Samoan homophone</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘bathe’</td>
<td>*kaukau</td>
<td>tā’ele</td>
<td>ta’ele ‘keel, bottom’</td>
</tr>
<tr>
<td>‘big’</td>
<td>*lasi, *nui</td>
<td>tele</td>
<td>tele ‘plane’ (V)</td>
</tr>
<tr>
<td>‘coral lime’</td>
<td>*lase</td>
<td>namu</td>
<td>namu ‘mosquitoes’</td>
</tr>
<tr>
<td>‘dog’</td>
<td>*kulii</td>
<td>maile</td>
<td>(laul) maile ‘Alyxia sp.’</td>
</tr>
<tr>
<td>‘front’</td>
<td>*muqa</td>
<td>luma</td>
<td>luma ‘shame’</td>
</tr>
<tr>
<td>‘full’</td>
<td>*fonu</td>
<td>tumu</td>
<td>tumutumu ‘top’</td>
</tr>
<tr>
<td>‘shark’</td>
<td>*magoo</td>
<td>malie</td>
<td>malie ‘agree’</td>
</tr>
</tbody>
</table>

(3) In the next group, the Samoan word is not a single morpheme, but appears (at least diachronically) to be polymorphemic. To varying degrees, the meanings of these new formations look like plausible literal or allusive descriptions for the meaning in question.

Table 3: Polymorphemic items

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
<th>Literal meaning of Samoan</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘centipede’</td>
<td>*molokau</td>
<td>atua-loa</td>
<td>‘long god’</td>
</tr>
<tr>
<td>‘forehead’</td>
<td>*laqe</td>
<td>mua-a-ulu</td>
<td>‘front of head’</td>
</tr>
<tr>
<td>‘fruit dove sp.’</td>
<td>*kulukulu</td>
<td>manu-tagi</td>
<td>‘weeping bird’</td>
</tr>
<tr>
<td>‘garden’</td>
<td>*maqala</td>
<td>fa’a-ta’</td>
<td>‘planting’</td>
</tr>
<tr>
<td>‘hawksbill turtle’</td>
<td>*kea</td>
<td>i’a una</td>
<td>‘tortoise-shell fish’</td>
</tr>
<tr>
<td>‘swamp hen’</td>
<td>*kalae</td>
<td>manu-ali’i</td>
<td>‘chiefly bird’</td>
</tr>
<tr>
<td>‘tongue’</td>
<td>*qalelo</td>
<td>laulau-faiva</td>
<td>‘calls out work’(?)</td>
</tr>
<tr>
<td>‘tree sp. (Vitex)’</td>
<td>*lala</td>
<td>namu-lega</td>
<td>‘smell of turmeric’(?)</td>
</tr>
<tr>
<td>‘turtle’</td>
<td>*fonu</td>
<td>lau-me</td>
<td>‘breadfruit leaf’</td>
</tr>
<tr>
<td>‘white tern’</td>
<td>*akiaki</td>
<td>manu-sina</td>
<td>‘white bird’</td>
</tr>
</tbody>
</table>

(4) A small number of replacements may be borrowings from other languages, such as Tongan.

Table 4: Possible borrowings

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
<th>Tongan</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘burning’</td>
<td>*wela</td>
<td>(mu)mū</td>
<td>mumū ‘warm oneself at fire’</td>
</tr>
<tr>
<td>‘go (singular)’</td>
<td>*fano</td>
<td>alu</td>
<td>‘alu ‘go (singular)’</td>
</tr>
<tr>
<td>‘go (plural)’</td>
<td>*oro</td>
<td>oo</td>
<td>oo ‘go (plural)’</td>
</tr>
</tbody>
</table>
Finally, in a few cases, the Samoan word has, as far as I know, neither external cognates nor homophones within Samoan.

**Table 5:** Samoan terms with no known external cognates or internal homophones

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘finished’</td>
<td>*qoti</td>
<td>‘uma</td>
</tr>
<tr>
<td>‘painful, hurt’</td>
<td>*mamahi</td>
<td>tīgā</td>
</tr>
<tr>
<td>‘run’</td>
<td>*tere, *lele</td>
<td>momo’e</td>
</tr>
<tr>
<td>‘sea, salt water’</td>
<td>*tahi</td>
<td>sami</td>
</tr>
<tr>
<td>‘shade’</td>
<td>*malu</td>
<td>paolo</td>
</tr>
<tr>
<td>‘white’</td>
<td>*tea</td>
<td>pa’e</td>
</tr>
</tbody>
</table>

The following examples might be regarded, not as real lexical replacement, but simply irregular sound changes. These changes are not part of the regular pattern of sound changes from PPn to Samoan, and in some cases (*k > m, *p > t) would be phonetically surprising in any language.

**Table 6:** Samoan terms with possible PPn sources (following irregular sound changes)

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘blow nose’</td>
<td>*fegu</td>
<td>fogi⁵</td>
</tr>
<tr>
<td>‘lick’</td>
<td>*qepo</td>
<td>eto⁶</td>
</tr>
<tr>
<td>‘navel’</td>
<td>*pito</td>
<td>pute</td>
</tr>
<tr>
<td>‘near’</td>
<td>*tata</td>
<td>lata</td>
</tr>
<tr>
<td>‘pattern’</td>
<td>*kanu</td>
<td>manu</td>
</tr>
<tr>
<td>‘skin, peel’ (V)</td>
<td>*fofole</td>
<td>fofoe</td>
</tr>
</tbody>
</table>

The meanings of most of the examples above fall within ‘basic vocabulary’ — not necessarily in the strict Swadeshian sense, but in a wider sense including locally salient flora and fauna and cultural items. For this reason we can reconstruct PPn forms for the meanings in question with some confidence. The Samoan words are innovations relative to PPn. In most cases they are unique to Samoan, though some are also found in Tokelauan. They are ‘replacements’ in the sense that they are a common everyday word for the given meaning. In some cases the PPn etymon is completely lost in Samoan; but in many others it survives as an alternative, in marginal or altered senses or restricted usages. Thus alongside *sami ‘sea, salt water’ we have *tai ‘tide’; while *umi means ‘long (object)’, *loa persists for ‘long (time)’. Samoan *leaga replaces PPn *saqa ‘bad’, but a reflex of the latter survives as *saa ‘sacred, forbidden’ (replacing PPn *tapu). A certain number of ‘replaced’ etyma survive in the vocabulary of respect (*’upu fa’aaloalo) (Milner 1961), thus reversing the normal pattern in which the ‘common’ terms are those that reflect PPn.

---

⁵ The phonetically modified form may reflect the influence of *sogi ‘smell, sniff’.
⁶ The change from labial to lingual consonant here might be seen as iconically motivated.
Table 7: Respect vocabulary

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Common</th>
<th>Respect</th>
<th>PPn</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘bathe’</td>
<td>tā’ele</td>
<td>‘au’au</td>
<td>*kaukau</td>
</tr>
<tr>
<td>‘die’ (non-human)</td>
<td>pē</td>
<td>matte</td>
<td>*mate</td>
</tr>
<tr>
<td>‘dog’</td>
<td>maile</td>
<td>‘uli’</td>
<td>*kulii</td>
</tr>
<tr>
<td>‘fear’</td>
<td>fefe</td>
<td>mata’u</td>
<td>*mataku</td>
</tr>
<tr>
<td>‘rain’</td>
<td>timu</td>
<td>ua</td>
<td>*quha</td>
</tr>
</tbody>
</table>

Undoubtedly some of the above replacements are instances of ‘normal’ lexical change (even if that means no more than changes for which no particular motivation is known). In many cases, however, there is evidence to suggest that patterns of homophony and avoidance have played a role. Consider first some examples where the apparently avoided word has an anatomical meaning which could be regarded as indecent. A striking example is the word for ‘tongue’ (PPn *qalelo). The expected Samoan reflex, alelo, exists, but is glossed by Pratt (1861) as ‘the tongue, a term of greatest abuse’. Milner (1966) glosses it as ‘eyes of a snake or eel, Adam’s apple (offensive)’. Exactly how alelo acquired these meanings and connotations remains to be determined. However, it is clear that laulaufaiva has completely supplanted it as the ordinary word for ‘tongue’. A combination of semantic and formal relation to *qalelo may also have caused the PPn verb for ‘stick out the tongue’, *faatelo, to be replaced by Samoan fa’aeto (fa’a- ‘CAUS’ + eto ‘lick’). Some further possible examples of indecency-avoidance are shown in Table 8.

Table 8: Replacements possibly motivated by indecency avoidance

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
<th>Source of replacement</th>
<th>Avoided word</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘fast’ (V)</td>
<td>*pali</td>
<td>‘anapogi’</td>
<td>‘last night’(?)</td>
<td>pali ‘pubes’</td>
</tr>
<tr>
<td>‘hide’ (V)</td>
<td>*fufuu</td>
<td>nanā</td>
<td>truncation(?)</td>
<td>fū ‘vagina’</td>
</tr>
<tr>
<td>‘navel’</td>
<td>*pito</td>
<td>pute</td>
<td>vowel changes</td>
<td>pitopito ‘anus’</td>
</tr>
<tr>
<td>‘ridge pole’</td>
<td>*taqufufu</td>
<td>‘au’au</td>
<td>*kau ‘stick’</td>
<td>fufu ‘onanism’</td>
</tr>
<tr>
<td>‘tree sp. (Vertex)’</td>
<td>*lala</td>
<td>namulega</td>
<td>‘smell of turmeric’</td>
<td>lalala ‘pudend. muliebr.’</td>
</tr>
</tbody>
</table>

Such patterns of homophony are much more widespread among the lexical replacements, but without any obvious anatomical indecency to be avoided. Here we need to consider an explanation in the phenomenon broadly known as word-tabooing. This practice is widespread in Oceanic languages in a variety of modes (Simons 1982). The form prevalent in Samoan was explained by the missionary linguist George Pratt (1861:89):

---

7 All avoided words and their glosses are from Pratt (1861). None of them appear in Milner (1966), whether because they were no longer in use, or because they were considered unprintable by those who commissioned the later dictionary.

8 Pudenda muliebría ‘woman’s private parts’. Here, as elsewhere in the dictionary, Pratt uses the common Victorian taboo device of switching to Latin for expressions presumably too indecent to be read by any but the educated.
A word is *tabooed* if it is a chief’s name, and may not be used by anyone in the presence of the chief holding the name as a title. Thus the name of the king of Manu’a is *moa*, therefore a fowl is a *manu*. The bat or *pe’a* is called *manulagi* where the name of the ruling chief is *Pe’a*. *Talo* is *fu’āuli* in Falealili, because the former is the name of a chief there. Similarly the vine or *fue* is *fau*, where one name is *Fue*. Similarly also the names of gods, were not allowed to be used by their worshippers as common words, e.g., *Vave* for which *ta’alise* was substituted in that district.

Pratt’s examples of local lexical replacements exemplify two of the types already illustrated: semantic shift and compound formation.

**Table 9:** Local lexical replacement

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>General Samoan</th>
<th>Local</th>
<th>Source of replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘bat’</td>
<td><em>peka</em></td>
<td><em>pe’a</em></td>
<td><em>manulagi</em></td>
<td>‘sky bird’</td>
</tr>
<tr>
<td>‘fowl’</td>
<td><em>moa</em></td>
<td><em>moa</em></td>
<td><em>manu</em></td>
<td>*manu ‘bird’</td>
</tr>
<tr>
<td>‘quickly’</td>
<td><em>wawe</em></td>
<td><em>vave</em></td>
<td><em>ta’alise</em></td>
<td>?</td>
</tr>
<tr>
<td>‘taro’</td>
<td><em>talo</em></td>
<td><em>talo</em></td>
<td><em>fua-uli</em></td>
<td>‘fruit of shoots’</td>
</tr>
<tr>
<td>‘vine’</td>
<td><em>fue</em></td>
<td><em>fue</em></td>
<td><em>fau</em></td>
<td>*fau ‘shore hibiscus’</td>
</tr>
</tbody>
</table>

The tabooing practice described by Pratt (1861), which hinges on personal names of chiefs and gods, might better be termed *name-avoidance*. The Tahitian name-avoidance system has been discussed in some detail (White 1968), though it ceased to operate during the 19th century; and there are brief allusions to such practices elsewhere in Polynesia (see Williams 1971:xxxvii for Maori). The most detailed description of a living name-avoidance system in Oceania has been provided by Keesing and Fifi’i (1969) for the Kwaio speakers of central Malaita, Solomon Islands. The **Kwaio** system involves avoidance of the names of significant ancestors, and usage of avoidance terms is restricted to local descent groups. Nevertheless, it is not unreasonable to suppose that the relation of the avoidance system to the formal structure of the language will be similar in principle to the documented Polynesian systems.
Table 10: Homophones and name avoidance

<table>
<thead>
<tr>
<th>Meaning</th>
<th>PPn</th>
<th>Samoan</th>
<th>Source of replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘turtle’</td>
<td>*fonu</td>
<td>laumei</td>
<td>‘breadfruit leaf’</td>
</tr>
<tr>
<td>‘full’</td>
<td>*fonu</td>
<td>tumu</td>
<td>?</td>
</tr>
<tr>
<td>‘deep’</td>
<td>*fofonu</td>
<td>loloto</td>
<td>*loto ‘inside, lagoon’</td>
</tr>
<tr>
<td>‘lagoon’</td>
<td>*loto</td>
<td>aloalo</td>
<td>*alo ‘front, lagoon side’</td>
</tr>
<tr>
<td>‘inside’</td>
<td>*loto</td>
<td>totonu</td>
<td>*loto + *tonu ‘right in the middle’</td>
</tr>
<tr>
<td>‘forehead’</td>
<td>*laqe</td>
<td>mūāulu</td>
<td>‘front of head’</td>
</tr>
<tr>
<td>‘swamp hen’</td>
<td>*kalae</td>
<td>manuali‘i</td>
<td>‘chiefly bird’</td>
</tr>
<tr>
<td>‘three’</td>
<td>*tolu</td>
<td>fia (rare)10</td>
<td>*fiha ‘few’</td>
</tr>
<tr>
<td>‘thick’</td>
<td>*matolu</td>
<td>mafia</td>
<td>as above</td>
</tr>
<tr>
<td>‘sew’</td>
<td>*tuitui</td>
<td>su’isui’</td>
<td>*suki ‘prick, stab’</td>
</tr>
<tr>
<td>‘candlenut (Aleurites)’</td>
<td>*tuitui</td>
<td>lama</td>
<td>*rama ‘torch’</td>
</tr>
<tr>
<td>‘white’</td>
<td>*tea</td>
<td>pa’e(pa’e)</td>
<td>?</td>
</tr>
<tr>
<td>‘fish sp. (Caranx)’</td>
<td>*malatea</td>
<td>malauli</td>
<td>*quli ‘black’</td>
</tr>
<tr>
<td>‘daytime’</td>
<td>*qahoeata</td>
<td>aauaui</td>
<td>as above</td>
</tr>
</tbody>
</table>

Keesing and Fifi‘i (1969:166–168) include a discussion and exemplification of mechanisms by which replacement forms are derived: (1) Normal semantic shift; (2) ‘Radical’ semantic shift; (3) Compounding and new morphological formations; (4) Borrowing; (5) Coining a ‘radically new term’; and (6) Phonological modification. (Apparently even a slight change in one vowel or consonant may suffice to satisfy the requirements of avoidance.) These devices are exemplified in the six groups of examples given above.

Pratt’s (1861) testimony shows that name-avoidance was still a living practice in the 19th century. (It may still have some currency, but I know of no modern studies on the subject.) But in all five cases he mentions, the ‘General Samoan’ words (reflecting PPn) survive as the basic words today, and only two of the local replacements are found in Milner’s (1966) dictionary (ta’alise ‘hurry’ as a near-synonym of vave, and fuāuli as a respect word for taro). Clearly only a small proportion of lexical avoidances would have had a permanent effect on the language. Given that the above examples do result from such avoidance, we may never be able to identify an actual chief or god whose name was the basis of a particular change.

Nevertheless, the evidence reviewed suggests that homophony and avoidance, both of anatomically indecent words and of chiefly or divine names, have played a significant role in the shaping of the modern Samoan vocabulary. In trying to clarify the relations among different lexical items even within a small range of meaning, I am struck once again by what a vast and complex system is a dynamically evolving lexicon.

Pratt (1861) simply lists *fia as a synonym of *tolu ‘three’. Milner (1966) marks it as obsolete, with the note ‘Occurs in ceremonial titles’, but also gives what appears to be a normal sentence example.
References


Language as fun and secret code: some play varieties in the Pacific

ROBERT EARLY

1 Introduction

The extensive multilingual repertoires of Pacific individuals and speech communities are well known. There is also a recognition that within any particular Pacific language, there may be significant geographical dialectal variation, and some Pacific languages also display highly developed systems of social dialect specialisation, such as the rank-based lexical variation in Tongan depending on whether one is speaking to the monarch, a member of the nobility or chiefly class, a commoner, or, in previous times, to a member of the slave class.

Another domain of language use which may entail the development of special varieties is that of play language, especially where the phonology of the standard language is manipulated in various ways to produce varieties which can be quite unintelligible to speakers of the standard form. These languages can therefore serve as markers of a special group identity, and as an effective means of keeping in-group communication exclusive and hidden from non-member participation. However, the highly systematic way in which these language varieties are formed does mean that the scrambling rule or code used to form them can often be broken or learned quite readily by other speakers.

Speakers who develop these special languages do so by responding to underlying perceptions about segments and syllables, and so an analysis of the manipulations they undertake can show up what some of these underlying perceptions must be.

This paper looks at some limited elicited data sets for several play languages in Bislama and Fijian, as well as another variety from Samoa which stretches the ‘play’ language designation somewhat, and gives a few initial observations regarding their function and form.

1 Sections of this paper were first delivered as a conference presentation at COOL6 (Sixth Conference on Oceanic Linguistics), Vila, Vanuatu, 4th–9th July 2004. Andrew Pawley, as always, showed great interest, and it is a great honour to have been one of his students, and to participate in this volume. Happily, the two Oceanic languages which possibly had Andy’s longest and most abiding interest, Samoan and Fijian, are well-represented in this paper. I would like to thank the speakers who discussed their speech varieties with me, and allowed themselves to be recorded: Kalisi Kauyaca and Litia Vuniduvu (Fiji), Maria, Evelini, Silia, and Alex (Samoan), and Jessica Stephens and Eslyn Kaltongga (Vanuatu). Thanks also to Jane Kanas and Wilford Gibson for assistance with capturing images on video and file. Video clips of most of the data used here can be accessed at http://plu.vanuatu.usp.ac.fj/Resources/.
2 Ludlings

The varieties that we are looking at here fit the technical description of ‘ludlings’, a term which was proposed by Laycock in 1965 (as reported in Laycock 1969:14, fn.36). This term, derived from Latin ludus ‘game’ or ludere ‘to play’ and lingua ‘language’, avoids the too specific term ‘pig latin’, and the too general term ‘play language’. Ludlings are one type of a larger class of special languages, or ‘sublanguages’ (to use a word from Laycock 1969:1), which includes baby talk, children’s languages, secret languages with lexical substitution, opposite talk, poetic language, call languages, whistle languages, disguised speech, and drum and trumpet signals. There are others as well, such as ‘T-ing in I’s’, or talking in initials, mentioned by Crystal (1997:59).

Laycock’s (1972) formal definition is as follows:

Ludlings are a category of special language which are ‘the result of a transformation or a series of transformations acting regularly on ordinary language text, with the intent of altering the form but not the content of the original message, for purposes of concealment or comic effect.’

A good listing with brief descriptions of many ludlings can be found in the online encyclopaedia, Wikipedia. The host languages mentioned, and the numbers of play languages which are formed from them, include Dutch (1), English (8, including Pig Latin, Cockney Rhyming Slang, and Gibberish), Farsi (1), Finnish (2), French (2), German (2), Hebrew (1), Hungarian (1), Cantonese (1), Indonesian (1), Italian (1), Japanese (1), Mandarin (1), Portuguese (2), Russian (2), Spanish (2), and Swedish (4).

Crystal (1997:59) mentions just a few in his discussion of secret languages, but provides a wider spread, referring to the more fully described play languages in English, French, Javanese, Thai, and Cuna (Panama). He describes them as ‘“languages” in which words are systematically altered, through the addition, subtraction, substitution, or transposition of sound’. Crystal also notes that some ludlings are manipulations based on spelling and alphabetical order, rather than being based on spoken language. The development and use of these languages that are going to be intelligible to only a small number of close associates is related by Crystal (1997:58) to the need for marking membership in a group, or to ensure secrecy for some reason, or simply as a pastime. This latter reason is stated rather informally, but it does capture some aspect of the human propensity to explore and appreciate various manifestations of linguistic creativity.

The most extensive bibliography of relevant materials appears to be an electronic source in LinguistList, Jones (1994a, 1994b). A later listing with information about various varieties of ‘secret signalization codes’ is Mahdi (1997). These show that some studies of these varieties for other non-European languages go back a long way (e.g. Conklin 1956 for Tagalog, and Aufinger 1948 for Madang, Papua New Guinea), and that other language families or languages where their existence has been noted include Amerindian, Arabic, Bakwiri (Africa), Bengali, Taiwanese, Burmese, Buin, Estonian, and others. Most discussions make reference to extensive studies by Bagemihl (1988, 1996) and Laycock’s (1972) typology of play languages.

The modifications usually involved in turning normal language into a ludling form are either reversals, or insertions, or a combination of both. The reversals are often at the syllable level, and may be of all or just one syllable (e.g. last syllable comes first). Backwards talk involves reversal at the level of segments. It is interesting to note that where the insertion is a consonant, it is more often /b/ or /p/ than any other consonant, and this is what we will see taking place in one of the Bislama varieties.
There are several cases of generalised vowel replacement reported: for Swedish, and for a language from Brazil called Lingua Do I (Language of i), where presumably the host language is Portuguese. Jones (1994a) also mentions an English ludling called Reetspeak, which is the speech of rat-like creatures in a science fiction book. Interestingly, all these ludlings use only /i/ as the replacer vowel, and this is exactly what we will see taking place in Fijian as well below.

An example of a ludling involving a combination of both reversal and insertion is classic English Pig Latin, whereby *sleep* becomes *eep-slay*, with the onset sent to the end of the word, and followed with the meaningless insertion *-ay*.

It is clear that such varieties are pervasive. Fromkin comments: ‘I don’t think anyone has ever studied a language community without finding one or more’ (Fromkin 1991). Crystal (1997:58) also comments on their ubiquity, and includes them within a larger categorisation of secret and hidden languages and codes. Laycock (1972:135) also considered that such forms of play language might prove to be ‘endemic in Austronesian languages, and can perhaps be expected to be widespread in non-Austronesian languages as well’.

Laycock also commented that his discoveries of these special language varieties made him realise how little was known ‘of forms of language other than the most mundane — or at any levels other than those necessary for grammars’ (1972:144). Perhaps he would be disappointed to find today that since making this observation, there has been very little further progress in the identification, description and analysis of such language varieties, in the Pacific region at least. David Gil (pers. comm.) has studied five ludlings in Malaysia, but there is not much else for Oceania.

One exception is an excellent unpublished research essay by a student at the University of the South Pacific, Padric Harm, which was unknown to me when collecting my own Fijian data. Harm’s paper (1998) describes syllable reversals at the word level, and vowel substitutions in Fijian, and further reference will be made to this later.

Some interesting work has been done analysing the status of the elements that are inserted into strings in ludlings. Sanders (1999, 2000) proposes an overall theory of intra-representational correspondences which actually unites three distinct morphological processes within one framework, based on the fact that all of them involve the insertion of semantically-empty morphemes. The three are *reduplication*, which involves the insertion of a RDP morpheme, a process of *truncation*, involving a morpheme TRUNC (involved in systematic shortening of, e.g. first names), and a reversal process for ludlings (Sanders does not consider infixal ludlings), involving a LUD morpheme. The similarities between these three are clear:

- they are all semantically empty morphemes which are constructed from the phonological material of the base
- the shape of the realisation conforms to prosodic units, especially the syllable
- despite being semantically empty, they do seem to convey some kind of iconic meaning: reduplication = repetition, continuance; truncation = diminutive form of nouns; ludling = unknown material, therefore secrecy or hiddenness (Sanders 2000:4).

While Sanders considers in another paper that this approach ‘brings ludlings in from the fringe of phonology’ and provides a ‘theoretical correlation to the frequency and
learnability of ludlings’, he also concedes that the analysis does not yet account for infixing ludlings or those which combine reversal with infixing (Sanders 1999:8–9).

Sanders has also introduced the term ‘ludlingant’ to refer to the substring of the host form which undergoes reversal (Sanders 2000:31), and presumably this would also be used to refer to the infixed string in infixing ludlings.

Others have considered that studies of ludlings can inform the study of language generally. For example, Prentice expressed the hope that ‘more extensive information on the distribution and nature of play languages may even permit an eventual reconstruction of a particular (play language) strategy, providing similarly valuable insights into the structure of an ancestral language’ (1982:154). This may turn out to be the case, but no-one has done it yet; however, as we will see, the way ludlings are structured can tell us something about the phonologies of the host languages.

3 Bislama

3.1 p language

The first ludling that I came across for Bislama is called lanwis p or p lanwis in Bislama, and p language when referred to in English. (Note that L1 influence results in many Bislama speakers not distinguishing /b/ and /p/, and so some will also use ‘b’ when referring to, or writing out, this language). The name clearly refers to the ludlingant, but does not describe it fully, which itself is an interesting observation. Speakers say that the language is formed by inserting /p/ in words, but in fact the insertion is of the form /pV/ where V is the same vowel as the preceding one in the host. The insertion follows each vowel, or, in other words:

\[ V_1 \rightarrow V_1pV_1 \]

It might be noted that vowel harmonisation is prevalent in the vernacular languages of Vanuatu (e.g. subject person prefixes in verbs in Lewo, Epi), and its effect can be seen in the following example, where:

\[ I kat fulap aranis blong Aniwa i stap long maket. \]

‘There are lots of oranges from Aniwa at the market.’

becomes:

\[ Ipi kapat fupulapap aranapanipis blopong Apanipiwapapi stapap lopong mapakepet. \]

Once interest is shown in these forms, it won’t be long before some speaker will volunteer that they, or one of their friends, are able to speak other ‘languages’ like this. Jessica was one such person, dobbed in by her friends who recognised her facility in speaking these varieties. She composed a small self-introduction to display her skills, which simply said:

---

2 The form aranapanipis here is irregular, as aparapanipis would be expected. This may be a performance error, with the speaker omitting to apply the transformation until the third syllable, then applying vowel harmony with \( \alpha \) as if it was the first or second syllable, and then attempting repair with the -nipis sequence. Unfortunately, this was noticed at a time when it was no longer possible to elicit a repetition, which, it is anticipated, would have included the regular expected formation.
Nem blong mi Jessica [Jesika], mi blong Maewo, mi skul long Lycée [Lise].
Long fyuja mi wantem kam wan tija.
‘My name is Jessica, I am from Maewo (Island), I (go to) school at the
Lycée. In the future I want to become a teacher.’

First of all she readily performed a p language version (omitting the second clause):

Nabamebem blobong mibi Jebešikabada, mibi skubah lobong libisebe. Lobong
fyubyujaba mibi waban wabam wabam tibijabap.

Note here that the form for nem ‘name’ is a curious mix of response to both Bislama
speech and English spelling. The na-ba string gives an /a/ pronunciation to the ‘a’ letter in
‘name’, and the me-be gives an articulation to the silent final ‘e’ in ‘name’, but then the
word still finishes with the original final consonant /m/. Jessica is a francophone (French-
educated) Bislama speaker, and so the fyu-byu string for fyuja indicates the retention of the
high front rounded vowel, and indicates that the yu sequence in Bislama is regarded as a
diphthong, a single vocalic unit.

3.2 lk language

Immediately after this, Jessica volunteered to give an ‘lk language’ version. The
ludlingant in this case is lvkV, again with the vowel harmony process at work, so that:

\[ V_1 \rightarrow V_1lV_1kV_1 \]

Nelekem blolokong miliki Jelekiselikikalaka, miliki skukulololokong
likiseleke. Lolokong fdukujalaka miliki walakatelelem kalakam walakan
ilikijalaka.

At the start of this one, Jessica struggled with deciding what to do with nem ‘name’. At
first she tried a strategy like p language, which came out as nalakamelelem, which
responds to English spelling again, but she wasn’t satisfied with this and changed it to
nelekem, which is indeed a regular formation. The form walakatelelem has lost the first
consonant of the nt cluster, as walakantelelem would be expected from wantem.

I also asked Jessica to write out her renditions as well, and it was clear that her
anglicised and gallicised spellings of Bislama exemplified the non-standardised Bislama
that is used by most ni-Vanuatu who have been taught English and French literacy, but
never exposed to the concept of a standardised spelling for Bislama. The transcriptions
given here are my own, but a further step in this study would be to analyse more carefully
the significance of spellings she does use, such as, in this case, words like socololl, lylykicéléké, fulukaturkur, tealecherlaka. It seems it was harder for her to maintain a
consistent output in writing than in oral performance.

3.3 z language

The next named variety that Jessica produced was ‘z language’. This is similar to p
language:

\[ V_1 \rightarrow V_1zV_1 \]

Nazemezem/Nezem blozo mizi jizesižikazap, mizi skuzul lozong ligizezi. Lozong
fuzutazar mizi wažantezem kazam wazan tižijaza.
Again she was in two minds about what to do with nem ‘name’, but finally settled on nezem. The form blozo is still a regular formation on blo, a frequent contraction of blong.

This ludling is of particular interest because it shows that the ludlingant need not be restricted to using segments from the inventory of the host language only, as /z/ is not a Bislama phoneme.

Jessica and her friends asked themselves if there were any more languages like this, and one reminded them of the s language. It fits the same pattern as the p language and z language:

\[ V_1 \rightarrow V_1sV_1 \]

Nasemesem/Nesem blosong misi Jesesisikasa, misi skusul losong lisise. Losong fusuutar misi wasantesem kasa wasan tisijasa.

3.4 Generalisation of the process in Bislama

The above four types are the only ones which were named and identified as being varieties which Jessica and her friends have used themselves. But by this stage some of the others were becoming somewhat blasé about Jessica’s performances, and one of them said that it’s really nothing special because actually anyone can make up any language they like at any time, using any sound they want, like ‘l’ or whatever. So Jessica took the hint, and proceeded to produce an ‘l’ language version on demand:

Nelem blolong mili Jellesilika, mili skulul lolong Lilisela. Lolong fulutur mili walatelem kalam walam tilijala.

Interestingly though, her performance here was not so systematic. For fyuja ‘future’ she ended up with fulutur, with the second half of the word as in French. Also, the wantem again dropped the nasal consonant to become walatelem instead of expected walantelem, and there seemed to be a rhyming effect in kalam walam for expected kalam walan.

3.5 Generalisation of the process to other languages

Although these modifications have been described as creating a variety of Bislama, Jessica readily volunteered that she can inject exactly the same modifications into the other languages she speaks. She was readily able to produce b language, lk language, z language and s language varieties for English and French. She does not speak a vernacular language, and some speakers who know the ludlings for Bislama and do speak vernaculars comment that it has not occurred to them to make ludlings out of the vernaculars. Perhaps this is because in many social contexts in Vanuatu, where each of the vernaculars are so small, and there are so many of them, they are already operating as markers of in-group identity, and as secret language varieties, for their speakers. Indeed, one of the explicit rationales expressed by many ni-Vanuatu of the need for the retention of vernacular languages is that without them, there would be no secret language variety that they could turn to when required. Underlying this is an assumption that there may well be a need to keep secrets from others who do not speak your vernacular, but not so likely with co-speakers, reinforcing the close linkage of language with identity in the Melanesian consciousness.

Here are transcriptions of the French original sentence, and its ludling versions (at this point, these are exactly as written out by Jessica, after she had spoken them):
**Language as fun and secret code: some play varieties in the Pacific**

Je m’appelle Jessica, je vais à l’école au Lycée. Dans le futur j’ai l’intention de devenir une professeur.

French p language (here using b):

Jebe m’abapebelle Jebebibicaba, jebe vaibus abai l’èbebole auabul lybycebe. Danbang lebe fubutubur jaibai l’inbintendentionbin debe debeveibenibur ubune probofesebuver.

French lk language:


This example shows an interesting transcriptional device which Jessica uses a lot, which is to write the base word in the French spelling that she is accustomed to, but to present the ludlingant modification in a more phonemic or Bislama spelling, as in l’èlekecolokole.

French z language:


In this one, the transcriptional insertion rule includes post-vocalic nasals:

\[ V_1(N) \rightarrow V_1(N)zV_1(N) \]

However, this is more likely to reflect Jessica’s intuitions that vowel-nasal combinations in French are manifested by a single nasalised vowel segment.

French s language:


As with the other French examples above, we can see that the liased l’ definite article is clearly regarded as a syllable initial consonant and part of the target for following insertion, as in l’esecosole.

### 3.6 Talking in initials

Another kind of secret language mentioned in the surveys is where words, usually only key lexical items, surface with just the initial letter of the word being pronounced. Some subjects report doing this in Bislama, and give the example of talking about smoking in the toilets at school, where:

Oli go smok long toelet. ‘They’ve gone to smoke in the toilets.’

becomes

Oli go s long toelet.
4 Fijian

Harm was already familiar with an infixal ludling from his own Fiji Hindi language, and gives the following example, where the ludlingant is $rfV$, and where the vowel regressively assimilates to the preceding vowel of the host word:

$$Tum kaha jata hai? \quad \text{Where are you going?}$$

$$\rightarrow Turfum karfaha jarfata harfai. \quad \text{(Harm 1998:2)}$$

This is interesting because in contrast to the French examples above, where VV sequences were regarded as diphthongs (hence $vais$ became $vailaikais$ in French $lk$ language or $vaisais$ in French $s$ language), the last word in this Fiji Hindi example splits the VV sequence, with the ludlingant inserted after the first V.

Harm also knew of an English ludling which was used by schoolgirls in Fiji which inserted $lVfV$:

$$I \text{ won’t take any more calls.}$$

$$\rightarrow Alafai wolofon’t telefaike elefenilife molofo collofol. \quad \text{(Harm 1998:2)}$$

The first word here reinforces what was noted above about the first vowel of VV sequences being a target by itself, and the last word suggests that in some words at least (not won’t here) the ludlingant picks up the first consonant in a cluster in the coda, to become $lVfVC$.

Whether there is any relationship between the $rfV$ insertion in Fiji Hindi and the $lVfV$ insertion in English here is unknown, but we will see later that the $lVfV$ ludlingant is one that I have recorded for Fijian, and this shows again, as we have seen for Bislama, how easily speakers can transfer these modifications across languages.

Later Harm became aware of similar kinds of modifications that were taking place in the Fijian spoken around him, so he investigated them more closely. The two phenomena he looked at were syllable reversal and vowel substitution.

Syllable reversal appears in selected words in utterances, and his subjects considered that such modifications began in very recent times, and in particular among the gay community, along with occurrences of opposite talk ($let’s stay = let’s go; my wallet is full = I’m broke$ etc.). Its use has extended now to the mostly female friends of gays, and then more widely to anyone who wants to say something that will unfailingly evoke laughter. It is used as a genuine secret talk on occasions as well, and as a euphemistic way of swearing.

The vowel substitution reported by Harm is reported to have been known for several generations, so much so that it has a name, being known as $vosa vaka-kiki$ (speaking in the manner of $kiki$). It involves changing every vowel that occurs into the vowel $/i/ \text{ (note that in Fijian, orthographic } g \text{ is a velar nasal } [ŋ], \text{ and } c \text{ and } q$, which appear below, represent a voiced interdental fricative $[ð]$ and prenasalised voiced velar stop $[ŋg]$ respectively):

$$O au sega ni kila. \rightarrow I ii sigi ni kili. \ ‘I don’t know.’ \quad \text{(Harm 1998:8)}$$

This indeed was the first Fijian ludling which I became aware of in listening to friends speaking in Fijian around me. My passive understanding of Fijian was enough for me to often pick up on what they were saying in normal speech, and so when they wanted to say something without me understanding, they reverted to the use of this secret variety. It was pervasive and could be sustained over a number of turns, although participants differed in
Language as fun and secret code: some play varieties in the Pacific

their facility with it, and it usually broke down into great mirth after a while. Also, once the code was cracked, and it was no longer secret, it wasn’t used anymore.

However, once interest was expressed in this, one 20-year-old speaker, Kalisi, volunteered firstly another three ludlings which she uses or has heard, and on another occasion, another two. They all fall into the same genre of secret languages, and are regarded as involving some kind of similar linguistic modification. However, while the one we have already noted involves a regular vowel change, the other five involve an infixal ludlingant and vowel harmony.

Kalisi selected a short self-introduction to demonstrate these varieties. After speaking the sentences in the standard Bauan dialect, she also transcribed them.

Na yaca-qu o Kalisi Kauyaca.
ART name-1s.POSS ART (firstname) (surname)
‘My name (is) Kalisi Kauyaca.’

Au vuli tiko e na koro-ni-vuli ni vuli vu-ni-wai.
1s study PRS PREP ART place-GEN-study GEN study base-GEN-water/medicine
‘I’m studying at the medical school (school of studying to be doctor).’

Au vuli vu-ni-wai ni bati.
1s study base-GEN-water GEN tooth
‘I’m studying to be a dentist (doctor of teeth).’

Au yabaki rua tiko.
1s year two PRS
‘I’m (in the) second year.’

4.1 Vowel change to i

The schema for this variety is straightforward:

\[ V \rightarrow i \]

The first of the sentences above appears in this form as:

\[ Ni yiciqi i Kilisi Kiici. \]

This shows the diphthong au appearing as just a single i, and the ya CV syllable dropping the semi-vowel onset and becoming just i.

4.2 Insertion of postvocalic s, jV, stV, kVr, and lVfV

The three further Fijian ludlings which Kalisi volunteered involve the following schemas:

\[ V_1 \rightarrow V_{1s} \]
\[ V_1 \rightarrow V_1jV_1 \]
\[ V_1 \rightarrow V_{1st}V_1 \]

Later, when telling friends about how she had been demonstrating these varieties, she was reminded by them of two more, which she knew of but hadn’t thought to mention earlier. They are produced using the following rules (the occurrence of phonetic [g] is marked here and below, to distinguish it from orthographic g which is [ŋ]):
The identity between the vowel which triggers the process, and the vowel in the insertion shows complete vowel harmony in all these ludlings. The following table gives a syllable by syllable listing of Kalisi’s first sentence as produced in each of the ludlings mentioned, beginning with the \( i \) vowel change variety and then the five other insertion ludlings.

<table>
<thead>
<tr>
<th></th>
<th>Na</th>
<th>ya</th>
<th>ca</th>
<th>qu</th>
<th>o</th>
<th>Ka</th>
<th>li</th>
<th>si</th>
<th>Ka</th>
<th>u</th>
<th>ya</th>
<th>ca</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i )</td>
<td>ni</td>
<td>yi</td>
<td>ci</td>
<td>qi</td>
<td>i</td>
<td>ki</td>
<td>li</td>
<td>si</td>
<td>ki</td>
<td>i</td>
<td>ci</td>
<td></td>
</tr>
<tr>
<td>( s )</td>
<td>naja</td>
<td>yaja</td>
<td>caja</td>
<td>[g]juju</td>
<td>ojo</td>
<td>kaja</td>
<td>liji</td>
<td>siji</td>
<td>kaja</td>
<td>uju</td>
<td>yaja</td>
<td>caja</td>
</tr>
<tr>
<td>( ja )</td>
<td>nasta</td>
<td>yasta</td>
<td>casta</td>
<td>qustu</td>
<td>osto</td>
<td>kasta</td>
<td>listi</td>
<td>sisti</td>
<td>kasta</td>
<td>ust</td>
<td>yasta</td>
<td>casta</td>
</tr>
<tr>
<td>( kar )</td>
<td>nakar</td>
<td>yakar</td>
<td>cakar</td>
<td>qukur</td>
<td>kokor</td>
<td>kakah</td>
<td>likir</td>
<td>sikir</td>
<td>kakah</td>
<td>ukur</td>
<td>yakar</td>
<td>cakar</td>
</tr>
<tr>
<td>( lafa )</td>
<td>nalafa</td>
<td>yalafa</td>
<td>calafa</td>
<td>qulufu</td>
<td>olofo</td>
<td>kala</td>
<td>lilili</td>
<td>silili</td>
<td>kala</td>
<td>ulufu</td>
<td>yalafa</td>
<td>calafa</td>
</tr>
</tbody>
</table>

Fuller transcriptions of all the sentences using the insertions are also given:

**Fijian \( i \) vowel change:**

\[Ni \ yici\ i \ Kilisi \ Kiici. \ I \ vili \ tiki \ ini \ kirinivili \ ni \ vili \ viniwi. \ I \ vili \ viniwi \ ni \ biti. \ I \ yibiki \ rii \ tiki.\]

**Fijian \( s \) insertion:**

\[Nas \ yascas \ casquos \ Kaslistis \ Kasusyascas. \ Asus \ vuslis \ tiskos \ esnas \ kosrosnisvuslis \ nis \ vuslis \ vusniswasas. \ Asus \ vuslis \ vusniswasas \ nis \ bastis. \ Asus \ yasbaskis \ rusas \ tiskos.\]

**Fijian \( jV \) insertion:**

\[Naja \ yajacaja\ [g]juju \ ojo \ Kajalijisiji \ Kajaujuyajacaja. \ Ajauju \ vujuliji \ tijikojo \ ejenaja \ kajoronorjivujuliji \ niji \ vujulijivujunijivajajiji. \ Ajauju \ vujuliji \ vujunijivajajiji \ niji \ bajatiji. \ Ajauju \ yajabakakaji \ rujauja.\]

**Fijian \( stV \) insertion:**

\[Nasta \ yastacastaqusto \ ostost \ Kastalistisisti \ Kastaustuyastacasta. \ Astaustu \ vustulisti \ tistikosto \ estenastu \ kosterostonistivustulisti \ nisti \ vustulisti \ vustunistiwastasti. \ Astaustu \ vustulisti \ vustunistiwastasti \ nisti \ bastatisti. \ Astaustu \ yastabastakisti \ rustuasta \ tistikosto.\]

**Fijian \( gVr \) insertion (all \( g \) here are \([g]\]):**

\[Nagar \ yagaracagarqurugur \ kgor \ Kagarligirsigir \ Kagaruguryagaracagar. \ Agarugur \ gogornerguvururligir \ nigi \ kgorrogornigirvugurligir \ nigi \ vugurligivugurnigirwagarigarigir. \ Agarugur \ vugurligir \ vugurnigirwagarigarigir \ nigi \ bagartigir. \ Agarugur \ yagarbagarkigir \ ruguragar \ tigirkogor.\]
Fijian $IVfV$ insertion:

\[
\text{Nalafa yalafacalafaquulufu olofo Kalafalilisilifi Kalafaulufiyalafacalafa.}
\]

\[
\text{Alafaulufu vulufulilifi tilifikolofo elefenalafa koleforolofonilisivulufulilifi nilifi vulufulilifi vulufunilisiwalafailifi. Alafaulufu vulufulilifi vulufunilisiwalafailifi nilifi balafatilifi tilifikolofo. Alafaulufu yalafabalafakilifi rulufualafa.}
\]

Some observations on the Fijian data are as follows:

- in the vowel replacement language, the diphthong is replaced by a single occurrence of the replacement vowel, however in the insertion languages, the diphthong is broken into its two segments, each of which is modified individually,

- the Fijian ludlings allow for the phonotactic conventions of the language to be flouted. Both the $s$ and $kVr$ languages involve having all syllables become closed syllables with $/s/$ and $/t/$ codas respectively, resulting in non-usual consonant clusters like $[st]$,

- feature constraints on the phonological inventory are also expanded, so that the sound $[g]$ occurs in one of the ludlingants, although it is not otherwise part of the phonemic inventory,

- in the $j$ language, Fijian orthographic $q$ [$\tilde{g}$] becomes velar plosive [$g$] in some cases.

It should be noted that a major lacuna in the data is the lack of information on what happens to long vowels in such modified speech varieties.

5 Samoan

For some time on various visits to Samoa I have heard people mention the special ‘language’ that was reported to be spoken as an in-language by members of the fa’aafafine community. It was clear that this involved some kind of manipulation of standard Samoan. Most people talking about this language will cite as a single example the syllable-reversal of leai ‘no’ to aile, and regard it as a somewhat stigmatised slang variety, highlighting the marginalisation component of social perceptions of the fa’aafafine community, which exists alongside a significant strain of tolerance and acceptance as well. The variety does serve as an identity marker for this social group, along with some accentual variations (e.g. aspiration and palatisation of $/t/$, rhoticisation of non-native $/t/$ in borrowed words, and some switching of $/l/$ to $/r/$), but its use outside of this group is certainly not unthinkable. Others may use this variety to mock or tease fa’aafafine, either in their presence or not, and its use is also found among other girls who hang around with fa’aafafine. The characterisation of these other ‘good-time girls’ as prostitutes is probably too harsh, but they are certainly socially marginalised too, and their use of this variety as an in-group marker may also reflect their recognition of this categorisation. They may also use this.

3 I hesitate to provide a translation for this term for those who are unfamiliar with it. Drawing on Besnier’s work, Worth (2001) discusses the unsuitability for the Samoan paradigm of terms like ‘gay’, ‘homosexual’, ‘transvestite’, ‘transgender’ as translation equivalents. The terms ‘queen’ or ‘drag queen’ seem to be used by the community for self-reference in English (Peteru 1996).
variety as a secret language in the presence of speakers of standard Samoan, who usually say that they are not able to understand what is being said.

Niko Besnier (pers. comm.) does not think that this variety has been reported on elsewhere, nor does he think that there is a similar variety amongst the equivalent fakaleitī community in Tonga. He adds that there has been some publication on ‘gay language’ in Britain which apparently involves some kinds of word play. There is some suggestion that the noi lay play language of Vietnamese, which involves swapping initial consonants of pairs of words, may have developed as ‘a special occupational language of (female) prostitutes’ but may now be ‘somehow identified with gay people’ (Proschan 1997), although this was disputed in later postings on the same discussion list. Certainly as noted earlier in this paper, this kind of syllable reversal does appear in the play language variety of gays in Fiji.

While the production strategy for insertion ludlings can usually be expressed as a simple schema, as for the Bislama and Fijian ludlings above, the same is not so true for reversal ludlings. Some simple examples of bisyllabic reversal in Samoan are:

\[
\begin{align*}
\text{tasi} \text{ ‘one’} & \rightarrow \text{sita} \\
\text{fasi} \text{ ‘hit’} & \rightarrow \text{sifa} \\
\text{fale} \text{ ‘house’} & \rightarrow \text{lefa} \\
\text{moe} \text{ ‘sleep’} & \rightarrow \text{emo} \\
\text{mea} \text{ ‘thing’} & \rightarrow \text{ame} \\
\text{‘ato} \text{ ‘bag’} & \rightarrow \text{to’a}
\end{align*}
\]

The first example quoted, of leai becoming aile, suggests that the final VV sequence, which in standard analyses of Samoan is two syllable peaks, is regarded as a diphthong, and we have a bisyllabic form \(S_1S_2\) reversing to \(S_2S_1\). Further examples of this, including the VV sequence \(ou\), are given (in Samoan, orthographic \(g\) is \([ŋ]\), ‘is the glottal stop \([?]\), and long vowels have macrons).

\[
\begin{align*}
\text{maile} \text{ ‘dog’} & \rightarrow \text{lemai} \\
\text{Malia} \text{ ‘person’s name’} & \rightarrow \text{Liama} \\
\text{Silia} \text{ ‘person’s name’} & \rightarrow \text{Liasi} \\
\text{tatou} \text{ ‘we (incl.)’} & \rightarrow \text{touta} \\
\text{tāua} \text{ ‘we (incl. dual)’} & \rightarrow \text{uatā} \\
\text{va’ai} \text{ ‘see’} & \rightarrow \text{‘aiva} \\
\text{alu} \text{ ‘go’} & \rightarrow \text{lu’a}
\end{align*}
\]

In various languages, glottal stop patterns as a marginal consonant, and more like some special feature of the following vowel. In the va’ai example above, it patterns as a normal C would be expected to pattern, but in V-initial alu it is inserted without motivation. In the two examples below, it occurs as part of a word-level schema which is not affected by syllable reversal. ‘

\[
\begin{align*}
\text{‘oe} \text{ ‘you (sg)’} & \rightarrow \text{‘eo} \\
\text{a’u ‘I’} & \rightarrow \text{‘u’a}
\end{align*}
\]

Within bimorphemic words, each morpheme is independently reversed.

\[
\begin{align*}
\text{fa’a-kali ‘CAUS-wait’} & \rightarrow \text{‘afa-liku}
\end{align*}
\]
However, there is also an example of just the consonants reversing, but the vowels continuing in their original position.

\[ \text{kilo-kilo ‘RDP-glance’ } \rightarrow \text{ liko-liko} \]

Some four-syllable forms where bimorphemic status is less clear involve a reordering of \( S_1S_2S_3S_4 \) to \( S_3S_4S_1S_2 \), as in:

\[ \text{ta‘avale ‘vehicle’ } \rightarrow \text{ valeta’a} \]

There is an example of a long vowel being reanalysed as containing a sequence of two like vowels. Someone called out \( ufuufu \) to someone passing by, and I had the misfortune to be inquisitive enough to find out that this is actually a reversal of the form \( fi\text{ifu} \) ‘masturbate’, being used in this case as an expletive.

There are also some apparently idiosyncratic changes, but they may reflect wider patterns that could be identified with more data.

- a reversal of \( CV_1V_2 \) to \( V_1CV_2 \), with glottal word-initially:
  - \( \text{kea ‘care (borrowing)’ } \rightarrow \text{ ‘eka} \)
- the use of the voiceless palatal affricate \([t]\):
  - \( \text{tama latititi ‘small child’ } \rightarrow \text{ jema titila’i} \)
- a kind of assimilation or consonant harmony, whereby \( C_1V_1C_2V \) becomes \( C_2V_1C_1V \):
  - \( \text{tatala ‘talk’ } \rightarrow \text{ lalata} \)

The reversal process also only applies to open classes of lexical items, and some closed grammatical classes like pronouns, but not to other grammatical particles. This suggests that speaker competence divides all words up into two classes, with pronouns being classified more like lexical items than functors, and the modification is only applied to the output from one of these classes at some stage of mental processing. Many of these grammatical particles are monosyllabic, like the article \( le \) and the preposition \( i \), and therefore not susceptible to syllable shift, but some, like tense-aspect markers \( olo’o ‘present continuous’, ‘ua ‘completive’ and -ia ‘transitive marker’ are not, and yet these particles are not modified at all. On the other hand, the directional particle \( mai \) becomes \( ina \).

Some of the forms noted above where these modifications were made were elicited in isolation, but many come from longer utterances over which the modifications were applied in all possible locations.

\[ \text{fia’ai ‘oe ‘are you hungry’ } \rightarrow \text{ ‘ifa’ia ‘eo} \]

\[ \text{(note that both fia and fai seem to become ifa)} \]

\[ e \text{ alu ‘oe i le ta‘avale ‘you’ll go in the vehicle’ } \rightarrow \text{ e lu’a ‘eo i le valeta’a} \]

\[ \text{leai kea a’u ia ‘oe ‘I don’t care about you’ } \rightarrow \text{ aile ‘eka u’a ia ‘eo} \]

\[ e \text{ alu tatou i le fale ‘let’s go home’ } \rightarrow \text{ lu’a totua i le lefa} \]

\[ \text{fai-tatala-ia ‘o ‘oe ‘you’re talkative’ } \rightarrow \text{ ifa-lalata-ia ‘o ‘eo} \]

\[ \text{tatala le ‘ato ‘open the bag’ } \rightarrow \text{ lalata le toa} \]

\[ \text{olo’o galue le tama ‘the guy is working’ } \rightarrow \text{ olo’o luégæ le mata} \]

\[ \text{kilokilo mai le kama ia ‘oe ‘the boy’s looking at you’ } \rightarrow \text{ likoliko ima le mata ia ‘eo} \]

Note from the final two that whether the host form is in the Samoan \( t \) \((tama) \) or \( k \) \((kama) \) social dialect, the reversal is always only \( mata \).
A short statement by Alex, a fa’afafine law student at the Emalus Campus of USP in Vanuatu, shows how the changes can persist through a whole discourse.

<table>
<thead>
<tr>
<th>O a’u o Alex.</th>
<th>O u’a o Alex.</th>
<th>My name is Alex.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ou te nofo i Apia.</td>
<td>Ou te fonó i Apia.</td>
<td>I live in Apia.</td>
</tr>
<tr>
<td>Na ou alu mai i Vanuatu i le tausaga le lua afe ma le tasi.</td>
<td>Na ou lu’a ima i Vanuatu i le sagatau le lua afe ma le sita.</td>
<td>I came to Vanuatu in 2001.</td>
</tr>
<tr>
<td>O le manaiaia o le à’oga.</td>
<td>O le naiamaia o le oga’ā.</td>
<td>I enjoy my studies.</td>
</tr>
<tr>
<td>O le teleia o tama ma teine Samoa i inei i Vanuatu.</td>
<td>O le leteia o mata ma neite Sajik i inei i Vanuatu.</td>
<td>There are not many Samoan men and women here in Vanuatu.</td>
</tr>
</tbody>
</table>

Alex recounted that he feels quite left behind when he goes back to Samoa after being away for study in Vanuatu, as his friends comment on how he has problems keeping up with conversations where the reversal language is being used. He sometimes makes mistakes, such as using finefa as a reversal for fafinese ‘girl’, and having to be told the correct form, which is nefifa. This suggests yet another reversal strategy for this variety of Samoan, which is not just a switching of first and last syllables, but an ordered reversal $S_1S_2S_3 \rightarrow S_3S_2S_1$, which is then an example of backward talk, at the syllable rather than the segmental level.

Clearly at first glance, there are a number of different strategies being applied in the Samoan case. Perhaps with more data an overarching phonological schema could be developed to account for all or most of these modifications, but if not, some of the idiosyncratic reversals will have to be regarded as lexicalised forms. Alex’s talk shows the occurrence of individual lexical forms, with the occurrence of the form Sājik for Sāmoa, which adopts a folk etymology for the name of the country, with -moa coming from moa ‘fowl, chick’, which is then borrowed in translation back into Samoan.

6 Functions

There are some strong similarities in the provenance, diffusion, and usage of the play varieties used in Bislama and in Fijian. Some of the key features which apply in almost every case are:

- the play languages are invented, perpetuated, and used almost entirely by young female speakers,
- they often arise in contexts like boarding schools where young females are grouped together,
- the main reason for using these languages is as a secret code which boys will not understand,
- they provide a vehicle for girls to talk and gossip about the very boys who are listening to what they say,
- the motivation for inventing a new modification is when boys become familiar enough with the code that they start to understand it,
• girls who have spoken these languages with each other when younger will continue to use these languages for the same purposes, and to perpetuate their bonding into teenage years and early adulthood,

• use of any of these varieties by a speaker constitutes ‘being funny’, and is usually accompanied by laughter and hilarity.

The Samoan case is somewhat different, and appears to be a genuine in-group secret code, although like all secret codes, it becomes susceptible to learning by others.

Alex reports that the fa'afafine reversal language is still widely used by the fa'afafine community and their circles of friends. It appears that it can be, and is, used at a sustained level in casual interactions: some of them are ‘real professionals at using it, and talk it 24/7’.

7 Conclusion

This paper has described some of the frontiers to which speakers may go in extending their repertoire of speech styles, in the area of secret or play languages. The ludlings produced by Jessica display her flexible and almost precocious articulateness in making various systematic modifications to the phonological form of all the languages she knows. The oral performances of Kalisi were truly impressive in their speed of articulation, and with the inserted material often being a larger part of the utterance, the real message was truly obscured and unable to be decoded by the uninitiated.

The occurrence of these modified speech forms in Bislama, a contact language, is also of interest. The development of a pidgin into a creole is often tracked along lines of phonological, lexical and syntactic expansion, and extended usage into a wider range of domains, including occurrence as a mother tongue for some speakers. The development of special speech styles and genres is not usually taken into account, but it seems that the elaboration of speech styles in Bislama to include the range of ludlings mentioned here does mark a further stage of development in its evolution into a fully functional human language.

In the festschrift prepared for Don Laycock (Dutton et al. 1992), Dutton mentions one of Laycock’s early papers (Laycock 1974) on dictionary making in minor languages of the Pacific, which was a pioneering discussion of the particular problems for lexicography thrown up by various features of some Pacific languages. Dutton laments that Laycock was not able to extend his contribution in this area (Dutton 1992:125), but fortunately there have been many advances in that area since by those who have come behind, including of course Andrew Pawley.

However, there are other areas of interest which Laycock popularised among the Pacific linguistic community but which have not been followed up to any great extent. For example, Michael Young notes that he and Laycock had agreed to work together on a project collecting and plotting the distribution of ‘Melanesian scatology and salacious stories’, but time never allowed for it to eventuate (Young 1992:662). Another area of Laycock’s interests which has not been developed much further is his early work on glossalalia, and the same is true of his pioneering studies of Pacific ludlings. His comment that the variants ‘are still frequently ignored by the descriptive linguist’ (1969:1) is still somewhat true, but hopefully this paper will serve to encourage those engaged in language documentation to catalogue these seemingly fringe areas of linguistic performance, which nevertheless constitute an intrinsic component of the linguistic competencies manipulated by many speakers of Pacific languages as they create the contexts in which they participate with their interlocuters.
References


Bagemihl, Bruce, 1988, Alternative phonologies and morphologies. PhD dissertation. University of British Columbia


—— 1972, Towards a typology of ludlings, or play languages. Linguistic Communications (Working Papers of the Linguistic Society of Australia) 6, 61–113.


1 Introduction

It is both a pleasure and an honour to be able to offer this tribute to Andrew Pawley, a colleague and friend with whom I have shared many fruitful hours of discussion. The recollection of our many discussions has enormously enhanced my personal appreciation of the exceptional range of Andrew Pawley’s interests. In a single conversation, Andrew is able to discuss topics as diverse as the coding of events in Kalam, the categorisation of proto-Oceanic sailing technology or the competence of commentators on cricket. From English across a world of Oceanic and Papuan languages, there is hardly a topic that he is unable to engage in and provide insights. As a consequence, for a Festschrift of this kind, there exists a range of topics on which one might write in tribute.

For this volume, I have chosen to consider a topic — the understanding of formulaic language — on which Andrew has written at some length beginning two decades ago and continuing to the present (Pawley 1985, 2007). This is a topic on which I am currently focusing much of my research and one on which I have benefited from continuing discussion with Andrew. It is therefore a topic on which I can try to offer an appropriate tribute.

2 Oral formulaic language: some illustrations by Rotinese poets

In a recent paper, ‘Developments in the study of formulaic language since 1970: a personal view’ (2007), Andrew has pointed to the remarkable prevalence of speech formulae in language and the pressing need to comprehend the variety of these formulae ranging from fixed words, traditional word pairs, conventional expressions, pragmatic phrases, idioms, collocations to collocates comprising lexical or phrasal units. In the paper, he also surveys some of the various ‘research traditions’ that have developed in relation to the study of particular forms of formulaic language.
Among these traditions, the study of the use of formulaic language in oral epic poetry, research begun by Milman Parry (1930) and continued by Alfred Lord (1960), has prompted perhaps the most extensive efforts across the widest spectrum of languages (see Foley 1985). With the enormous proliferation of this research, many different research directions have emerged. For some literary scholars, formulaic language has been used as evidence to argue for an oral tradition underlying the production of ancient literary texts; for others, dealing with living poetic traditions, formulaic language has been seen as a means of sustaining coherent performance, allowing poets both considerable oral fluency and a capacity for substantial recall.

As Alan Rumsey (2001) has noted in an important recent paper, Parry’s original definition of a formula was ‘a group of words which is regularly employed under the same metrical conditions to express a given essential idea’ (1930:80). This definition only fits certain forms of oral poetry. The variety of the world’s oral poetry defies simple summary. Not all such poetry operates under strict ‘metrical conditions’; other compositional constraints may apply. Hence, as research on ‘oral formulaic language’ has increased, the definition of what is ‘formulaic’ has become elusive and needs to be carefully set forth for particular traditions of oral composition. More generally there is the question as to what precisely is ‘formulaic’ for a particular poet, or for a particular speech community, or across related speech communities. In this brief paper, I would like to consider some of these issues by way of illustration.

Since 1965, I have been recording oral poets, known as manahelo ‘those who chant’, from the island of Roti in eastern Indonesia. These recordings have resulted in a substantial corpus of ritual texts. The island of Roti was once divided into seventeen small domains, each of which claimed to be socially and culturally distinctive. Reflecting this social divergence, the language of the island consists of a chain of dialects whose ends are not mutually intelligible. To date, I have recorded mainly in the dialect of the central domain of Termanu and the south-western dialect of the domain of Thie, though I am now engaged in recording dialects across the entire island.

The corpus of the texts which I have gathered deals with the origins of the cultural artefacts of the Rotinese: of fire and cooking, of rice and millet, of the house and its designs, of bridewealth including water buffalo, and of the tools for building, for weaving, for dyeing and for spinning. The corpus also includes a large number of mortuary chants as well as prayers, sermons and parts of a local Christian liturgy which is composed in a traditional mode. Over several decades, I have published a considerable number of papers on individual texts (see bibliography), but I have never published comparisons among these texts. It is only by way of such comparisons that one can appreciate the formulaic aspects of this oral poetry.

In 1965, I began recording a particular chant (bini) known as Suti Solo do Bina Bane. Over time, I have recorded this chant from numerous oral poets, particularly in the domain of Termanu where I have done much of my research. Besides recording in the dialect of Termanu, I have recorded this chant in several other dialects and have been particularly interested in recording the same poet reciting the same chant at different stages in his career. As a consequence, my collection of Suti Solo do Bina Bane texts is itself substantial and continues to grow with each new recording session.

In this paper, I wish to compare the use of oral formulaic language by five different poets, each reciting what may be regarded as the same passage in the chant Suti Solo do Bina Bane. For one of these poets, I provide recitations of this passage at an interval of
several years to illustrate continuity in personal composition. Four of these poets come from the domain of Termanu; one comes from the domain of Ringgou, whose dialect is substantially different from that of Termanu. This offers an illustration of the use of oral formulaic language across different speech communities.

3 SutiSolo do Bina Bane as an oral composition

*Suti Solo do Bina Bane* recounts the journey of two shells cast ashore by the sea. These shells, whose dual name is *Suti Solo//Bina Bane*, are taken up in the scoop-nets of women fishing along the coast. Once on land, these shells continually voice their feelings of displacement and appeal for permanent companionship. In response, the women urge the shells to shift from one symbolic location to another. Thus they are told, for example, to find their place with the ‘Syrup Vat and the Rice Basket’, then with the ‘Millet’s Grain and Ears of Maize’, then with the ‘Tree’s Shade and Lontar’s Shadow’ or with the ‘Forest Cuckoo and River Watercock’. As each of these sites is vulnerable to change and can only be a temporary resting place, the shells continue to plead for permanence. Their metaphoric journey is a quest for an enduring place of rest.

Originally, I suspect, this chant was recited to reveal the origins of two different kinds of shell: the one used for holding dyes (*Suti Solo*) and the other used as the base for the spinning whorl (*Bina Bane*). However, as these rituals for dyeing and weaving have ceased to be performed, the chant has been transformed into a mortuary chant for ‘orphans’ — a category that can be applied to all human beings in conditions of dependency (see Fox 1988c). Therefore, depending on the intention of the poet, the chant can be given various endings. The shells may either be fashioned into implements or they return to the sea.

All origin chants may once have formed part of a single narrative structure that recounted relations between the Sun and Moon and the Lords of the Ocean and Sea (Fox 1997). Key passages in most versions of *Suti Solo do Bina Bane* that I have recorded betray the connections to this larger epic structure by alluding to events in the realm of the sea that caused the shells to be cast forth on land.

The excerpts that I have selected from each of the poets’ compositions vary in length from thirteen to nineteen lines. Each composition consists of a dialogue between the shells and the women who scooped them up from the sea. The women urge *Suti Solo//Bina Bane* to find a place with the Rice Basket and Syrup Vat. *Suti Solo//Bina Bane* reply that they will do so, but they worry that when the Rice Basket and Syrup Vat are emptied they will no longer have a place to remain.

Here it is important to note that the women who are named in this dialogue as well as the two shells are each conceived of as single dual-named beings. Hence the third person singular is generally used rather than any plural. Before presenting these compositions, it is essential to note the conventions of canonical parallelism that apply to them and the research tradition that has developed in the study of this form of poetic composition.

4 Canonical parallelism in relation to an oral formulaic

All compositions in Rotinese poetry (or ‘ritual language’ as I have termed these poetic compositions in many of my publications) are characterised by a strict lexical pairing. Apart from a small number of unpaired forms — pronouns, connectives, ‘prepositional’ and a few other invariant elements, all lexical terms have at least one pair. Thus, in formal terminology, each semantic element must form part of a ‘dyadic set’. In composition
dyadic sets produce parallel lines whose overwhelmingly most common poetic form is the couplet, though other serial arrangements of lines are entirely acceptable and are often considered as evidence of a greater mastery of the language. The elements that compose any particular dyadic set should, in their parallel lines (or occasionally in the two halves of a single line), correspond exactly in position and as far as possible in morphological structure.

Parallelism is a pervasive feature of much of the world’s oral literature. The linguist Roman Jakobson, who contributed greatly to the study of parallelism, described this study as opening ‘the double door linking the fields of linguistics and anthropology’ (see Fox 1977).

In Roman Jakobson’s terminology, the required lexical pairing of semantic elements and the network of associations that underlie this pairing represents a canonical ordering of the paradigmatic or ‘metaphoric pole’ of language (Jakobson and Halle 1956:76–82), but Rotinese ritual language is also remarkably well ordered along the syntagmatic, or what Jakobson called the ‘metonymic pole’ of language. Phrases and lines in these oral compositions are frequently composed of recognisable formulae that, because of the strict requirements of parallelism, become redoubled in parallel formulae. Thus, a considerable portion of Rotinese ritual language consists of couplets and even longer sequences that are formulaic in both a paradigmatic and a syntagmatic sense.

The term ‘parallelism’ derives from the 18th century investigations of Bishop Robert Lowth on the fixed semantic correspondences (parallelismus membrorum) found in the Old Testament (Lowth 1753, 1758). Since Lowth’s initial research, a large biblical literature has been devoted to parallelism. This research was later stimulated by the discovery of Canaanite (Ugaritic) texts that relied on a closely related corpus of parallel terms.

This ‘canonical’ coupling of words has been identified in oral compositions in numerous other languages of the world. The Kalevala is one of the most frequently cited examples of canonical parallelism. Various scholars have noted and discussed similar forms of parallelism for Ostyak and Vogul folk poetry, for Hungarian, for Mongolian, and for Turkic. Researchers have established parallelism as the first principle of Nahuatl and of Mayan poetry, including works such as the Popul Vuh. It is also found in the poetry of contemporary Maya groups, in Cuna folk literature and in Quechua poetry dating back to the time of the Incas.

Early on in the 19th century, scholars pointed to parallelism in Chinese poetry. Since then, parallelism has been noted in a variety of linguistic forms: in early written documents, in the rhyme-prose of the Han period, in ‘parallel prose’, in love songs, in proverbs and in popular poetry. Similar usages have been observed in Tibetan, as well as among the Thulung Rai of east Nepal, the Sadar of Jaspur and the Toda of South India. In Southeast Asia, the use of parallelism has been documented in compositions among the Kachin-, Kmhmu-, and Khmer-speakers, and in written and oral compositions in Vietnamese. Parallelism has also been observed in Australian languages (Strehlow 1971) as well as in the Highlands of New Guinea (Rumsey 1995, 2001).

A large textual literature on parallelism can also be found within the Austronesian-speaking world where extensive traditions of parallelism in poetry and ritual languages have been documented for the Rhade of Vietnam, the peoples of Nias, the Batak, a variety of Dayak groups throughout Kalimantan, among Bugis and Toraja groups in Sulawesi and for numerous populations in eastern Indonesia where vibrant traditions of oral composition in parallelism persist: Sumba, Flores, Savu, Rote and Timor. (See Fox (1977) for detailed
documentation on the variety of the world’s traditions of parallelism, and Fox (1988a and 2005) for detailed summaries of the Austronesian traditions of parallelism.)


Within a large corpus of textual materials the importance of these formulae becomes increasingly evident. While retaining the required pairing of words, poets add individual grammatical embellishments to distinguish their usage from that of other poets. Thus, individual poets develop their own personal ‘style’ in relation to certain ‘standard’ forms of their dialect area. The interaction of the formulaic features of ritual language with the rules of parallel composition creates further complexity. One question to be addressed is how these formulae and the pairs underlying them vary among the speech communities of different dialect areas, especially since a high proportion of synonymous dyadic sets are composed of a word from the local dialect and that of some neighbouring dialect.

5 Suti Solo do Bina Bane: the 1st poet: Stephanus Adulanu

This first excerpt — the Syrup Vat and Rice Basket sequence — consists of 19 lines from a poem of some 297 total lines. I recorded this poem from one of the senior poets of the domain of Termanu, Stephanus Adulanu, in 1966. Stephanus Adulanu was the head of the clan Meno. He was known as ‘Old Meno’ and held the ritual position of Head of the Earth (Dae Langak). His version of Suti Solo do Bina Bane is perhaps the finest and certainly the most extended version of this poem that I have gathered in Termanu.

Apart from personal names (Pedu Hange//Nggeti Seti) which themselves form sets, the poem is composed of ten dyadic sets: (1) kokolak//dede’ak ‘to speak/to talk’, (2) inak//fetok ‘woman/girl’, (3) eki//hika ‘to scream/to laugh’, (4) setele//mata-dale ‘shriekingly/gaily’, (5) tua bou//neka hade ‘syrup vat/rice basket’, (6) malole//mandak ‘good/proper’, (7) bou//soka ‘vat/sack’, (8) (lama-)-kako// (lama-)-lua ‘to overflow/to run over’, (9) fude//bafa ‘froth/mouth’, (10) totono//lulunu ‘to overturn/to roll up’ (partially reduplicated form of tono//lunu).

Of interest is the compound set, tua bou//neka hade ‘syrup vat/rice basket’. The components of this compound can and do occur on their own in other sets. Here, however, they appear to form a ‘personified’ set. Thus, in a literary sense, they suggest a ‘living entity’ who can serve as a potential partner of the shells. (Hence, in my translation of the poems, I have capitalised them to suggest their implied personhood.)

Passage: Poem I

‘Na Bina au o se
Ma Suti au o se
Fo au kokolak o se
Ma au dede’ak o se?’
Boe ma inak-ka Nggeti Seti
Ma fetok-ka Pedu Hange nae:
‘Te eki setele henin
Ma hika mata-dale henin na
‘Then I, Bina, with whom will I be
And I, Suti, with whom will I be
With whom will I talk
And with whom will I speak?’
The woman Nggeti Seti
And the girl Pedu Hange, says:
‘(If) they scream with a shriek at losing you
And laugh gaily at losing you,
Suti mo Tua Bou
Ma Bina mo Neka Hade.’
Boe ina nae:
‘Oo malole-la so
Ma mandak-kala so.
Te leo bou lamakako fude
Ma soka lamalua bafa
Fo bou lo totonon
Ma soka no lulunun
Na Suti au o se
Ma Bina au o se?’

Then Suti, go with Syrup Vat
And Bina, go with the Rice Basket.’
Then she says:
‘Oh, these things are good
And these things are proper.
But if the vats overflow with froth
And the sacks run over at the mouth
So that the vats must be overturned
And the sacks must be rolled up,
Then I, Suti, with whom will I be
And I, Bina, with whom will I be?’

This excerpt provides a basis to begin the consideration of a succession of other versions of this same passage composed by different oral poets.

6 Suti Solo do Bina Bane: the 2nd poet: Eli Pellondou

This second excerpt consists of 23 lines from a chant of a total of 210 lines. I recorded it from the poet, Eli Pelondou who was more commonly known by the nickname, Seu Ba’i. He was a proud, soft-spoken man of clan Dou Danga and was particularly close to Stephanus Adulanu. I first met him on a visit to Old Meno’s home in Ola Lain and it appeared to me that he had informally apprenticed himself to Old Meno. His version of Suti Solo do Bina Bane reflects, I believe, this influence.

This excerpt relies on six of 10 the dyadic sets used in Old Meno’s composition: (1) kokolak//dede’ak, (2) tua bou//neka hade, (3) malole//mandak, (4) lulunu//totono, (5) (lama-)kako//((lama-)lua, (6) bafa//fude, as well as two other dyadic sets not used in the Meno version: (7) (nama-)tani//((nasa-)kedu ‘to cry//to sob’, and (8) sama//deta ‘like//as’.

In this passage, Seu Ba’i recites one line for which he fails to provide a complement. Were this line to exist, the appropriate complement for the verb (masa-)lai ‘to rest, lay down’ would have been (manga-)tu ‘to sit’.

Pedu Hange//Nggeti Seti is the name of the woman who scoops up the shells. As is clear in the context of the longer poem, she is the speaker in this passage who tells Suti Solo//Bina Bane to ‘Go with the Syrup Vat//Go with the Rice Basket’. The woman, Lole Holu//Lua Bafa, is referred to as the ideal partner whom Suti Solo//Bina Bane longs to find.

Given that dyadic sets can either be treated as singulars or plurals, in any single poem, poets can move back and forth from singular to plural. In general, however, they tend to use singular verbs for dual chant characters. Thus, for the most part, Pedu Hange//Nggiti Seti or Suti Solo//Bina Bane are treated as single (singular) figures.

Passage: Poem II

Boe te Suti neu namatani
Ma Bina neu nasakedu,
Nasa-kedu Lole Holu
Ma nama-tani Lua Bafa.
Boe te ana dede’ak no Suti
Ma ana kokolak no Bina, lae:

But Suti begins to cry
And Bina begins to sob,
Sobs for Lole Holu
And cries for Lua Bafa.
So she speaks with Suti
And she talks with Bina, saying:
‘Mo tua bou
Ma mo neka hade
Fo masalai tua bou.
(Line missing)
Boe te Bina neu kokolak
Ma Suti neu dede’ak, nae:
‘Au u o tua bou
Ma au [u] o neka hade.
De malole ndia so
Do mandak ndia so.
Te neka lamakako bafa
Fo soka lo lulunun
Ma tua lamalua fude
Fo bou lo totonon,
Au dede’ak o se
Ma au kokolak o se
Sama leo Lua Bafa
Ma deta leo Lole Holu?’

It is particularly interesting to note that despite the fact that Seu Ba’i’s passage relies on many of the same dyadic sets as Old Meno’s passage, only two lines in these passages are the same. It is thus instructive to compare the way in which each poet expresses what might be considered the same ‘formulaic’ lines.

Old Meno gives the following two lines:

‘Oo malole-la so
Ma mandak-kala so.’

Seu Ba’i, on the other hand, renders this same formula as follows:

‘De malole ndia so
Do mandak ndia so.’

Old Meno then goes on to recite the following lines:

‘Te leo bou lamakako fude
Ma soka lamaluam bafa
Fo bou lo totonon
Ma soka no lulunun …’

In the last two of these four lines, Old Meno does what expert poets do frequently (to the maddening frustration of the translator): he creates a kind of contrastive dyadic set by coupling a plural form (lo) with a singular form (no).

Seu Ba’i renders these lines in a slightly different fashion. Where vat/sack (bou/soka) form a set in relation to the verbs (lama-kako/lama-lua) in Old Meno’s composition, Seu Ba’i uses the set, basket/syrup (neka/tua) with these verbs. He does, however, maintain a plural agreement throughout.
James J. Fox

‘Te neka lamakako bafa
Fo soka lo lulunun
Ma tua lamalua fude
Fo bou lo totonon…’
‘But if the basket overflows at the mouth
So that the sack must be rolled up
And the syrup runs over with froth
So that the vat must be overturned…’

Technically, soka lo lulunun/bou lo totonon are the only lines that both poets share yet even this seeming shared similarity has been altered by Old Meno’s use of a singular and a plural form.

7 Suti Solo do Bina Bane: the 3rd poet: Mikael Pelondou

Some years later, in 1985, on a brief visit to Roti, I was told of the death of Seu Ba’i. Among the group who came to tell me of his death was a clan cousin of his from Dou Danga, Mikael Pelondou. As far as I have been able to determine, the two men had the same great-grandfather. They had lived in close proximity to one another and referred to each other as ‘elder’ and ‘younger’. Mikael agreed to record for me his version of Suti Solo do Bina Bane. This passage is taken from that version.

It is an excerpt of 15 lines from a composition that runs to only 101 lines. Although shorter than the two previous excerpts, it contains many of the same sets. Mikael uses: (1) kokolak//dede’ak, (2) inak//fetok, (3) tua bou//neka hade, and (4) fude//bafo, but instead of (lama-)kako//lama-lua he uses (5) (lama-)kako//lama-solo which has much the same meaning. The only new set is a commonly used set, (6) (na-)tane//na-nosi ‘to ask/to request’. Also in Mikael Pelondou’s composition, the name, Pedu Hange//Nggeti Seti, is given as Pedu Hange//Suti Seti. Although this may seem only a minor difference, it is over the names of chant characters that poets have their greatest arguments.

Passage: Poem III

Boe te natane ma nanosi. But he still asks and requests.
‘Bina dede’ak no se ‘With whom will Bina speak
Ma Suti kokolak no se?’ And with whom will Suti talk?’
Boe te inak leo Pedu Hange So the woman like Pedu Hange
Ma fetok le Suti Seti, nae: And the girl like Suti Seti, says:
‘Dede’ak mo neka hade ‘Speak with the Rice Basket
Ma kokolak mo tua bou.’ And talk with the Syrup Vat.’
Boe te Suti natane But Suti asks
Ma Bina nanosi, de nae: And Bina requests, saying:
‘Au dede’ak o tua bou ‘I speak with the Syrup Vat
Ma au kokolak o neka hade And I talk with the Rice Basket
Te neka lamakako bafo, But if the basket overflows at the mouth
Na au dede’ak o se Then with whom will I speak
Ma tua lamasolo fude, And if the syrup runs over with froth
Na au kokolak o se?’ Then with whom will I speak?’
8 Suti Solo do Bina Bane: the 4th poet: Petrus (Pe’u) Malesi

Of all the poets of Termanu, I recorded the most material from Pe’u Malesi. He was a frequent visitor to my home at Ufa Len and probably the most willing of all poets in Termanu to seek to have his compositions recorded. (Malesi is the ‘poet’ in the film, Water of Words (1983), and appears as well in the film Spear and Sword (1988d). For a photograph of Malesi chanting at a ceremony in honour of Old Meno, see Fox (1989b:521).)

Like Seu Ba’i, Pe’u Malesi was a member of clan Dou Danga, but the two men were rivals. Seu Ba’i, in particular, would challenge Pe’u Malesi’s knowledge of the names of chant characters, and in fact, as is evident in Malesi’s telling of Suti Solo do Bina Bane, the names of the key women who scoop the shells from the sea are entirely different.

Due to the closeness of our relationship, I had the opportunity to record various versions of Suti Solo do Bina Bane as recited by Pe’u Malesi. Here I include two such versions, the one recorded in 1973 and the other in 1977. These short excerpts, which are remarkably similar, may be compared to one another and to the excerpts of the other poets of Termanu, Old Meno and Seu Ba’i.

Pe’u Malesi’s first version of the Syrup Vat and Rice Basket passage consists of only 14 lines from a composition of some 222 lines.

One immediately obvious difference in Malesi’s version is the name of the woman chant-character who carries on the dialogue with Suti Solo/Bina Bane. In place of Pedu Hange/Nggeti Seti, Malesi names this woman as Sama Dai/Kuku Nou. Another notable difference is that Malesi reverses the order of the compound phrase, tua bou ma neka hade. He recites this as bou tua ma neka hade. In terms of the ordering of dyadic sets, this would seem to be a preferable order: bou ‘vat’ thus forms a set with neka ‘basket’ and tua ‘lontar syrup’ with hade ‘rice’. Whereas this would appear to make logical sense and would not be rejected in performance, Malesi may in fact be tampering with an established idiom in Termanu. Ande Ruy, from the relatively distant dialect region of Ringgou with no connection with poets from Termanu, uses the same form as Malesi: bou tua ma neka hade.2

Malesi’s passage shares several common dyadic sets with the version by Old Meno or Seu Ba’i: (1) inak//fetok, (2) kokolak//dedea’ak, (3) malole//mandak, (4) tono//lunu. He also uses the dyadic set (5) (anga-)tu//(asa-)lai ‘sit//lay down’, a set that Seu Ba’i left incomplete in his composition. Malesi also uses another common set, (6) lole halan//selu dasin, which is a frequently used expression meaning ‘to speak, to reply’ and embellishes this with the set, lele//doko-doe: ‘encouragingly//coaxingly’.

Passage: Poem IV

| Inak Kia Sama Dai | The woman, Sama Dai |
| Ma fetok kia Kuku Nou | And the girl, Kuku Nou |
| Ana lole lele halan | She lifts her words encouragingly |
| Ma selu doko-doe dasin, nae: | And raises her voice coaxingly, saying: |
| ‘Mu no bou tua | ‘Go with the Syrup Vat |
| Ma mu mo neka hade.’ | And go with the Rice basket.’ |
| Bina Bane kokolak | Bina Bane speaks |
| Ma Suti Solo dede’ak ma nae: | And Suti Solo replies and says: |

---

2 In the domain of Landu, which historically has been a major producer of millet but not of rice, this set becomes ‘syrup vat and millet basket’: (bou tua ma fati bete).
‘Malole la so
Ma mandak kala so.
Bou tua na tono
Ma neka hade lulunu
Na au asalai o se
Ma au angatu o se?’

‘That would be good
And that would be proper.
[But if] the Syrup Vat is overturned
And the Rice Basket is rolled up
Then with whom will I rest
And with whom will I sit?’

This next excerpt, which consists of 16 lines, was recorded in 1977 roughly four years after the earlier passage but is remarkably similar to that passage. There are only a few differences. Thus, for example, the set lele//doko-doe is not used with lole halan//selu dasin and instead of the tono//lunu set, Malesi uses another set with a similar sense: heok//keko ‘to turn/to shift’. He concludes with a refrain that uses the set, (nama-) tani//(nasa-)kedu, ‘to cry/to sob’, which was used by Seu Ba’i in his composition. Malesi also pairs another verb for speaking (na-fada) with the term, kokolak which normally forms a set with dedea’k. In performance terms, this would be considered acceptable but nonetheless a flaw in composition.

**Passage: Poem V**

Boe ma inaka Kuku No’u
Ma fetoka Sama Dai
Lole halana neu
Ma selu dasi na neu ma nae:
‘Meu mo neka hade
Ma meu mo bou tuana.’
Boe ma Suti Solo nafada
Ma Bina Bane kokolak ma nae:
‘Ah, malole la so
Ma mandakala so.
[Te] bou tua la heok
Ma neka hade la keko
Na au asalai o se
Ma au angatu’u o se?’
Suti bei namatani,
Ma Bina bei nasakedu.

So the woman, Kuku No’u
And the girl, Sama Dai
Raises her voice
And lifts her speech and says:
‘Go with the Rice Basket
And go with the Syrup Vat.’
Then Suti Solo talks
And Bina Bane speaks and says:
‘Ah, that would be good
And that would be proper.
But if the Syrup Vats turn
And if the Rice Baskets shift
Then with whom shall I rest
And with whom shall I sit?’
Suti continues to cry,
Bina continues to sob.

9 Suti Solo do Bina Bane: the 5th poet: Ande Ruy from Ringgou

The final example in this series consists of a passage of 20 lines from a composition of 184 lines by the poet Ande Ruy, from the east Rotinese domain of Ringgou which has its own distinct dialect of Rotinese — a dialect that is different from that of Termanu but close enough to be intelligible. Some of the sound changes evident in this passage are:

<table>
<thead>
<tr>
<th>Termanu</th>
<th>Ringgou</th>
</tr>
</thead>
<tbody>
<tr>
<td>ng</td>
<td>k</td>
</tr>
<tr>
<td>medial k</td>
<td>–</td>
</tr>
<tr>
<td>l/nd</td>
<td>r</td>
</tr>
</tbody>
</table>
Despite these differences, it is possible to recognise the shared inheritance of common dyadic sets. Ande Ruy uses five sets that also occur in the passages of the poets from Termanu. These sets are: (1) inak//fetok, (2) lolo hara//selu dasi as well as (3) dasi//hala, (4) nea hade//bou tua, and (5) hade//tua. Ande Ruy also uses a number of other sets that are common in the speech community of Termanu: (6) fai//ledo: ‘day//sun’, (7) na//ria: ‘there//at that place’, (8) iku//leo: ‘land//clan’, (9) rui//sau: to scoop//scrape’, (10) tama//tesa ‘to be together/to be one’, (11) te‘i//dale ‘stomach//inside’. Although I have not recorded the set, sasau//kokola (reduplicated form of the verbs sau//kola), it would appear to be an acceptable variant of the more commonly used set, ndui//sau in Termanu (or Ruy//sau as it occurs in this excerpt). It also is worth noting that the compound form, tesa te‘i//tama dale (lit. ‘stomachs as one’//’hearts together’) used by Ande Ruy, is not an expression that I have recorded in Termanu. However, although not common in Termanu, it was immediately recognisable by Termanu-speakers who heard Ande Ruy and was considered poetically attractive.

In short, all of the sets in this particular passage can be considered as forming part of a shared linguistic heritage with speakers of the Termanu dialect.

Passage: Poem VI

<table>
<thead>
<tr>
<th>Line</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boe ma ina Oli Masi</td>
<td>Then the woman, Oli Masi</td>
</tr>
<tr>
<td>Ma feto a Bisa Oli</td>
<td>And the girl, Bisa Oli</td>
</tr>
<tr>
<td>Nadas neu Suti Solo</td>
<td>Spoke to Suti Solo</td>
</tr>
<tr>
<td>Ma nahara neu Bina Bane:</td>
<td>And said to Bina Bane:</td>
</tr>
<tr>
<td>‘iku fo mo nea hade ma</td>
<td>‘Your place is with the Rice basket</td>
</tr>
<tr>
<td>Ma leo fo mo bou tua.’</td>
<td>And your clan is with the Syrup Vat.’</td>
</tr>
<tr>
<td>Tehu Suti Solo lolo haran</td>
<td>But Suti Solo raised his words</td>
</tr>
<tr>
<td>Ma Bina Bane selu dasin:</td>
<td>And Bina Bane lifted his words:</td>
</tr>
<tr>
<td>‘Ami iku fo mo nea hade</td>
<td>‘Our place is with the rice basket</td>
</tr>
<tr>
<td>Ma ami leo fo mo bou tua, tebe!</td>
<td>And our clan is with the syrup vat, indeed!</td>
</tr>
<tr>
<td>Tehu fai esa nai na</td>
<td>But on some day</td>
</tr>
<tr>
<td>Ma ledo esa nai ria,</td>
<td>And at some time</td>
</tr>
<tr>
<td>Nea sasau hade,</td>
<td>They continually scrape out rice,</td>
</tr>
<tr>
<td>Sau heni nea hade</td>
<td>They will scrape the rice basket clean</td>
</tr>
<tr>
<td>Ma rui kokola tuan</td>
<td>And they continually scoop syrup</td>
</tr>
<tr>
<td>Rui heni bou tua.</td>
<td>They will scoop the syrup vat clean.</td>
</tr>
<tr>
<td>Na ami iku fo mo be a</td>
<td>Then with whom will our place be</td>
</tr>
<tr>
<td>Ma ami leo fo mo be a.</td>
<td>And with whom will our clan be</td>
</tr>
<tr>
<td>Te [bei] ta tesa tei</td>
<td>This does not make us one</td>
</tr>
<tr>
<td>Ma bei ta tama dale.’</td>
<td>And not yet join us together.’</td>
</tr>
</tbody>
</table>

10 The dyadic resources of the five poets

Table 1 lists all the dyadic sets used by the five poets. The excerpts from the poets of Termanu are similar enough in composition to share many sets together; Ande Ruy’s composition introduces seven sets not used by the Termanu poets. Yet all of these sets are of common occurrence in Termanu. Thus if a larger comparison were made, the seeming
differences in Ande Ruy’s composition — at least in the use of these particular sets — would diminish, if not disappear.

In various publications (Fox 1974, for example), I have argued that Rotinese ritual language surmounts dialect differences and is broadly intelligible across the entire island of Roti. One way in which this is done is by the use of variant dialect terms to form synonymous sets. Certainly the comparison of this passage from Ande Ruy with the excerpts by other poets from Termanu would appear to support this argument. However, while accurate in showing similarities between these speech communities, Ande Ruy’s excerpted passage is too short to illustrate some of the differences that do indeed occur. Providing a more precise indication of the differences in ritual language across dialect boundaries involves work that is currently in progress to record a reasonably large corpus of materials from these different dialects.

Table 1: Dyadic sets used in all poems

<table>
<thead>
<tr>
<th>Dyadic Sets</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. kokolak//dede’ak</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>2. inak//fetok</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>3. eki//hika</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. setele//mata-dale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>5. tua bou//neka (nea) hade</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>6. bou tua//neka hade</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. malole//mandak</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>8. bou//soka</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. lama-kako//lama-lua</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. lama-kako//lama-solo</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. fude//bafa (bafo)</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. totono//lulu nu</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. sama//deta</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. nama-tani//nama-kesu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>15. nama-tani//na-nosi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>16. lele//doko-doe</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. halan//dasin</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. nasa-lai//na-tu</td>
<td>(x)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. lole//selu</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. na-fada//kokolak</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. heo//keko</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. fai//ledo</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. ria//na</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. iku//leo</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. liai//sau</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. sau//kola</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. tesa//tama</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. te’i//dale</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10 Conclusion

In his pioneering study, Milman Parry was able to identify a particular ‘technique of oral verse-making’. His initial research has given rise, as Andrew Pawley has noted, to a distinct but particularly important research tradition. In this tradition, ‘formulas may show special word order, enabling a word sequence to be adapted to the metrical requirements of a half line of verse ...’ (Pawley 2007:6). This kind of formulaic is, as Pawley has recognised, a ‘substitution system’. He defines this substitution system as ‘a group of formulas which show lexical substitutions expressing the same basic structure and idea, or which express the same basic idea with a varying number of syllables, enabling the poet to meet a range of different metric conditions’ (Pawley 2007:6).

Strict canonical parallelism offers techniques of oral composition different from those based on regular metrical strictures. These techniques, strictly speaking, do not involve substitution as in Parry’s epic tradition nor are they constrained by metric requirements. Instead any line of verse calls forth its complement. All lexical elements in one line should pair with partner elements in a complementary line. Such compositions are ideally suited for chorus performance, in which a poet announces an initial line and the complementary line is provided by collective response, a mode of ritual performance that was once common on Roti as elsewhere in eastern Indonesia. To be effective, the knowledge of paired terms — what I call ‘dyadic sets’ — must be widely shared within a speech community.

This paper provides an illustration of just how widely these dyadic sets are indeed shared, not just among poets within a particular speech community but across different speech communities. This illustration gives a sense of the stability and continuity of this ritual language as a distinctive cultural heritage among the Rotinese — and its potential effectiveness in maintaining continuing oral-based memory.

References


—— 1983, Water of words: a cultural ecology of an eastern Indonesian island. (With Timothy and Patsy Asch.) 16 mm colour. 30 minutes. Watertown, Mass: Documentary Educational Resources.


—— 1988d, Spear and sword: a payment of bridewealth. (With Timothy and Patsy Asch.) 16 mm colour. 25 minutes. Watertown, Mass: Documentary Educational Resources.


Hoskins, Janet, 1988, Etiquette in Kodi spirit communication: the lips told to pronounce, the mouths told to speak. In J.J. Fox, ed. To speak in pairs: essays on the ritual languages of eastern Indonesia, 29–63. Cambridge: Cambridge University Press.


Exploring oral formulaic language: a five poet analysis


—— 1758, Isaiah X-XI. Translated as Isaiah, 1834, Boston.


New Zealand Māori plant names for teatree: manuka (Leptospermum scoparium) and kanuka (Kunzea ericoides), Myrtaceae

RHYS GARDNER

In the mid-1980s Ralph Bulmer asked me to look over his botanical specimens from Papua New Guinea’s Kaironk Valley, where for more than twenty years he, Bruce Biggs and Andy Pawley had been collecting all manner of things under the direction of their endlessly knowledgeable Kalam informants. Until then, working at the Auckland War Memorial Museum herbarium, I had steered clear of those specimens, their carefully-tagged Kalam names so unique as to give no identification clues at all. In this willing but unprepared kind of way I came to Andy’s attention, as he took over Ralph’s projects and set about getting scientific plant names for the forthcoming ‘animals book’ (Majnep and Bulmer 2007). I never became competent in Kalam orthography but I did gain an interest in ethnobiology. So, when the memorial volume for Bulmer appeared I found much of interest there, in particular Biggs’ now-classic review on the origins of New Zealand Māori plant-names (Biggs 1991).¹

When Andy subsequently took me to Fiji’s Waya Island, his second dictionary-making area of interest, I saw for the first time the small native tree Decaspermum fruticosum. It belongs to the same family as the two New Zealand teatree species and has the same oily-scented, narrow leaves and starry white-petalled flowers. Its names on Waya, and in Fiji at large, are nuga or naganuqa. In Tonga the species is known as nukonuka, in Samoa as nu’ana’u’a and in Tahiti (apparently) as anu’u’a. Clearly here was the ancestral referent for the manuka and kanuka of the New Zealand Māori.

At this point the story becomes one of several strands, which together might or might not bear the weight of an amateur linguist. (Or, as Andy might say, ‘He pureirei whakamatuatanga’!).² The so-called facts are these.

¹ I am grateful to Peter de Lange and Wendy Pond for information about names for teatree in northern New Zealand. The opportunities offered to me over the years by Andy and Medina Pawley (the ‘Pawley Foundation’) have been very much appreciated.
² ‘I rest for a moment on a swamp-sedge’, a metaphor for what the physicists call unstable equilibrium, i.e. ‘I’d better keep moving’!
Firstly, the cognates as above were not mentioned by Biggs (1991). That might just be an oversight, but remarkably, *Decaspermum fruticosum* has a gap in its Pacific Islands distribution — it is not known from the Cook Islands, despite there being plenty of suitable habitat there, at least on Rarotonga. Nor is *nuka* (as a plant name) part of any Cook Islands tale, proverb or saying as far as I know. That is, the putative ancestral home of the New Zealand Māori (today) contains neither the plant nor the name.

Secondly, it is not widely appreciated that the New Zealand Māori, at least in the northern part of the country, had other names for the two teatrees. Banks and Solander on Cook’s first voyage collected both taxa, but hardly appreciated their distinctness, and of the two Māori names they obtained for these plants the only intelligible one is *manuka*. It was some time later that collectors and travellers found that in northern New Zealand *Leptospermum scoparium* was sometimes called *kahikatoa*, and *Kunzea ericoides*, *ra(u)wiri*. Throughout the rest of the country, it seems that both teatrees were grouped together by Māori and early Europeans travellers simply as *manuka* — see the entries for the above names in Orsman (1997).

So it is tempting to conclude that at least one group of New Zealand Māori originated from a place where *Decaspermum fruticosum* was unknown.

The third part of this story is at least as remarkable. Island-biogeography theory predicts that after dispersing to new lands some taxa will barely be able to establish viable populations, while others, free of their homeland pests and diseases, will thrive until more strongly competitive species arrive. On the human time scale this process is only just perceptible (although collectors are always on the look-out for new arrivals that are not simply weeds of the moment but authentic members of the regional flora). The *nuka* story here may owe something to this process of establishment and decline. That is, in the Auckland War Memorial Museum there is a collection from Rarotonga obtained in 1899 by its very able director and botanist Thomas Cheeseman. It undoubtedly represents *Leptospermum scoparium*. This is the only time the species has ever been found outside New Zealand or Australia. Was it introduced deliberately by some homesick expatriate, or was it the last of its line, the ancient Cook Islands *nuka*? Alas, Cheeseman recorded no such name for his specimen, nor was he able to get information as to whether it represented a native plant or not. But if archaeologists find a layer of manuka brush and capsules in some Cook Islands dwelling-layer of a thousand years ago the story presented here might at last get a tidy ending.

References


Some thoughts on change in the Samoan language

CLUNY MACPHERSON

Pacific languages are undergoing transformation as globalising forces permeate Pacific societies ever more rapidly and deeply. This paper focuses on some sociological dimensions of this process of language change and in doing so, seeks to reflect and to emulate Andrew Pawley’s remarkable talent for contextualising the study of language. Andrew’s interest in the social context of language was not surprising given that early in his career he taught in the Anthropology Department at the University of Auckland with social anthropologists such as the late Ralph Piddington and Bruce Biggs who were also accomplished linguists. The message that the transformation of language is always a symptom, or consequence, of social and economic processes which are at work in a society, has stayed with me since I was first a student in Andrew’s Pacific linguistics classes over 40 years ago. This idea provides the framework for this discussion which is an attempt to locate some of the sociological forces which lie behind these changes rather than a technical account of linguistic shift. Knowing Andrew as I do, I am confident he will overlook the shortcomings of this attempt to reflect on lessons I learned in his classes all of those years ago.

The paper starts from the premise that change in Pacific languages is occurring ever more rapidly in those societies which are experiencing steadily intensifying contact with the forces of globalisation: commerce, tourism and migration. The shifts are not random: patterns in the character of change can be related to shifts in the social context in which a language is used. It is also argued here that globalisation does not invariably lead to the declining use of Oceanic languages with relatively small numbers of speakers. In fact, as the Samoan case illustrates, globalisation can produce both expansion and contraction in the language. This paper argues that Samoan has expanded as it has encountered new realities which the language has had to embody, and has contracted as opportunities to hear and use the language have declined, and as its users have chosen to use only parts of the language and for purposes other than communication. It further argues that after a long period of expansion, the Samoan language does face challenges to its long term survival.

The earlier part of this globalisation process resulted in a series of ‘growth spurts’ in the language as new people and technologies passed through, and settled in, Samoa and new words and, more importantly, concepts entered the language. More recently, contact with new peoples, ideologies and technologies, and competition from alternative languages in new sites has resulted in the shrinkage of Samoan. This is particularly evident in some
diasporic communities, where people are speaking versions of Samoan which have been transformed, truncated and in some places where Samoan has been displaced completely. While this shrinkage occurred initially in expatriate enclaves, the beginnings of this process are now also being seen in urban Samoa.

The earliest ‘globalisation’

To focus on the most recent phase of globalisation is, however, to overlook the possibility that Samoan has been changing for very much longer. The need for the first adoptions, of both concepts and words to signify them, into Samoan almost certainly began in prehistoric times with contact with non-Samoans who periodically visited Samoa from the Tongan and Fijian archipelagos, and occasionally from Uvea and Futuna, to trade, colonise and, later, to seek spouses. The difficulty of tracking the pace and character of incorporation during this period is that no means of identifying these changes with any precision is readily available. One might, however, speculate that, since visitors from the Fijian and Tongan islands came from similar physical environments and exploited similar marine and terrestrial ecosystems with parallel technologies, the numbers of adoptions connected with flora, fauna and technology might have been limited. One might also argue that early in this period of contact, the groups which visited each other had separated only relatively recently, and lived within similar social organisational arrangements and shared similar worldviews and languages. The differentiation which occurred as the frequency of voyaging declined later in this period (Howe 2007), was incremental, and was limited by the absence of ‘new’ material and social phenomena which languages had to incorporate.

Later globalisation

More profound changes began with intermittent and, later, permanent contact with a group of non-Oceanic people, who brought with them a new worldview, lifestyle and languages (Gunson 1978). Unlike early visitors from elsewhere in Oceania, these later visitors worldviews and lifestyles had been formed in very different cultural, political and economic milieux, in which different technologies were employed to exploit different ecosystems to different ends. Initially at least, Samoans did not have to confront, or accommodate, the entire cultural repertoires of these strangers. The early missionary visitors, driven by a desire to save the souls of the Samoans, were primarily concerned with the dissemination of parts of that repertoire which related to religion. Changes in other areas, it was supposed, would follow from the transformation of the religious sphere. While religion was only a part of the cultural repertoire, it was a fundamental one and connected central elements of the Samoan worldview and lifestyle.

The new religious model which the missionaries proposed defined a new, singular God and defined existing Samoan Gods out of existence; created a new set of intermediaries between humans and the new God and defined old intermediaries out of existence; transformed the character of relations between both humans and God and between groups of humans. The new, divinely ordained set of values which underpinned the new worldview mandated new lifestyle prescriptions and transformed a series of social institutions and practices in Samoan society. The new worldview redefined the nature and location of the afterlife, the relationship between people’s worldly conduct and their
destination in the afterlife. But significant parts of the new cultural repertoire could be incorporated by extending the use of ideas and words already in the Samoan language.

This solution could not work, however, in every case, and this became particularly apparent when it became necessary to translate the Bible into Samoan, a task which required new letters, words, and terms for ideas, practices, and names for places and people. By 1876, the missionary lexicographer, Reverend George Pratt, listed some 160 foreign words ‘in use among the natives’. The words listed came from English, Greek, Hebrew, Latin, Tahitian, Tongan and Rarotongan (Pratt 1911:103–104). These included terms for new items of fauna, such as aeto from Greek for the eagle and solofana from Tahitian for horse; flora, arasi from Hebrew for cedar and paina from English for pine; new technologies, pepa from English for paper and paeolo from English for barrel; minerals, ario from Latin for silver and teio from Greek for sulphur; social and religious practices, fa’aipoipo from Rarotonga for marrying and peritome from Greek for circumcising; units of time such as asolulu from Tahitian for Wednesday and all of the months from English and units of volume such as kalone for gallon and kupita for cubit from English.

These changes occurred alongside another series of endogenous shifts which Pratt also documented for several Oceanic languages. He noted that,

Some Polynesian tribes have recently changed the pronunciation of one or two consonants. The natives of Niue have changed the t into ts. Fifteen years ago only a portion of the younger people made the t into ts before i: as lautsi for lautii … In Hawaii they have changed the t into k, and ng into n. Thus tangata has become kanaka. Samoans are doing the same thing at the present time, to the great injury of the language. (Pratt 1911, Preface)

It was not difficult to simply borrow English terms into Samoan as the proliferation of Samoan versions of biblical names for people and places and terms proved. Indeed the Samoan alphabet was extended for this purpose but the use of Samoan versions of biblical names, such as Solomona, came to signify the hope that the recipient would show the same personal qualities as the biblical heroes. The creation of a series of new words using new sounds from an augmented alphabet could, however, only ever be part of the solution to the need to augment the Samoan worldview.

Missionaries faced a much greater challenge in explaining, and contextualising, the animals, objects, measures, polities, states, and physical phenomena which new words signified. These were not part of either the Samoan social or physical worlds, and neither their form, nor their metaphorical significance, was immediately apparent to Samoans. Soaring like eagles makes little sense without the concept of an eagle. To embrace many of the ideas in the Bible which are expressed as similes or metaphors, Samoans needed to be able to visualise camels and asps, gold and silver, cubits and bushels, Rome and Lebanon, deserts, devastating floods and plagues of insects. To overcome this problem missions published comprehensive manuals in the Samoan language such as Thomas Powell’s O le Tala i Tino o Tagata ma Mea Ola Ese‘ese e iai fo‘i o tala i manu ua ta‘ua i le Tusi Pa‘ia (A Manual of the Anatomy of Humans and Various Animals, with explanations of animals mentioned in the Holy Bible) (Powell 1886). These manuals, which were intended to allow Samoan teachers and missionaries to use these terms and concepts with authority, filled a series of gaps for Samoan lay readers of the Bible and allowed them to use these new ideas with authority. At the same time they extended the Samoan language in fundamental ways.
The presence of missionaries in Samoa, and periodic assurances that their evangelical activity was producing an increasingly stable and peaceful Samoan society (Moyle 1984:16), led almost inevitably to more general settlement of Europeans from Britain, the USA, Germany, France and Australia in search of cheap land and labour from which to create their fortunes (Gilson 1970). Their presence, in turn, was a major force for linguistic shift because to secure these commodities, they had first to commodify and to create exchange values for both of these things.

Land which, for Samoans, had only use value, became available through kinship and could not be alienated, had to be converted into a commodity which could be ‘measured’ precisely, given an exchange value, valued, and permanently alienated in a transaction involving a new portable medium of exchange. The ‘nature’ and ‘value’ of labour, which was traditionally ‘exchanged’ with kinship matrixes and used ultimately in the creation of sociopolitical capital, had also to be transformed and redefined as a commodity which could be sold in measurable units of time, to unrelated individuals for cash. The sale of labour to create a benefit which could be expropriated by an individual was a significant departure from the ‘traditional’ outcomes of the labour process.

The emergence of markets for land and labour, which had not been commodified but which were fundamental to profitable commerce, was only possible as a new vocabulary emerged to define and signify means of measuring, marking and designating areas of land; means of designating units of time in which labour could be sold, and the forms of agreement around the ‘sale’ of labour. These also required new forms of ‘contract’ which were enforceable by persons other than those directly involved, and the concepts of state and national law.

Transacting exchanges in these new markets depended also on the creation, and acceptance, of the idea of a new, generalised medium of exchange which used as its units of currency, items which had no obvious intrinsic value, to replace the well understood ‘currencies’ and ‘exchange rates’ employed in barter and ceremonial exchanges. Gold and silver coinage and paper notes had no obvious intrinsic value in a society which had not had access to any of these and no obvious means of establishing either their absolute ‘value’ or ‘equivalence’. How did one establish the value of a rectangular piece of paper, printed in an unintelligible foreign language, and its relationship to a series of small circular metal coins whose size bore no obvious, direct relation to their value? How was one to establish the equivalence of the different forms of coins and paper currencies involved in transactions between various people speaking different languages?

The creation of markets for labour and land, and the adoption of a new exchange medium, required the gradual adoption of a series of terms related to the sale of commodities which could not be sold or exchanged before. A series of new measures of weight, area and of time emerged to deal with this eventuality. Over time, a language emerged to provide these things. In some cases this language used Samoan words which had similar meanings [land/fanua; boundary/tuaoi]. These are similar but are not identical since land designated as ‘land’ in English was a measured commodity the value of which could be established by supply and demand and which could be disposed of by sale in a market. The term fanua in Samoan designated a family’s indivisible estate in common, the value of which was established by history and which could not be disposed of other than by gifting. In other cases, Samoan words were given broader or additional meanings [maps/fa’afamua; surveyor/fuafamua] and in other cases, non-Samoan words were (phonologically) adapted [acre/eka; foot/futu; surveying pins/pini] and became part of Samoan.
Along with these transformations came another which was fundamental to the establishment of capitalism, and involved the redefinition of the nature, and measurement, of wealth and of the purposes of economic activity. It involved validation of the idea that economic enterprise could be driven by accumulation of profit by individuals, which was re-invested in activities which further increased its owner’s personal wealth. Investment decisions in this model were driven primarily by rates of return on capital and the benefits to the individual. This represented a significant departure from the Samoan view that capital’s value was only enhanced in the redistribution process, and in which investment and re-investment decisions were driven by the need to disburse accumulated physical capital to increase the sociopolitical capital of groups rather than individuals.

The next source of expansion of the Samoan worldview, and the language in which it was embodied, came from the applications of new technologies to new crops to produce new by-products which were transformed into commodities, which Samoans had not seen or used, and which were sold in places to which Samoans had never been. These technologies came, in part from settler and, in part from missionary activity. Thus, a series of new specialisations connected with the establishment of commercial agriculture gave rise to further expansion of the language.

The new commercial settlers were in many cases anxious to take advantage of lower land and labour prices and find crops which could be sold on global markets at a profit (Gilson 1970). This almost invariably meant the introduction of new varieties of plants and trees, to produce new by-products, which were processed by new technologies. The management of introduced plants and their crops was controlled initially by the settlers and understood by a small number of Samoan labourers who worked for them. Samoans, who controlled the majority of the land, chose selectively to engage in some of these forms of agriculture where they were forced or when it suited them to do so. This trend, encouraged by missionary support for the adoption of new crops and technologies, ensured that the use of new concepts and terms was not confined to the small numbers of Samoans who provided labour to Europeans producers of these commodities, and found its way into more general use.

The introduction of cotton production in the mid-1860s was one such sequence. A shortage of cotton, caused by interruption to American production by the Civil War, led to a rapid increase in price and opportunities for growers in the Samoas. The introduction of a new crop, to produce a new textile, by a new technology led again to the expansion of Samoan to incorporate the language necessary to engage in this trade. Samoans, faced with drought and a shortage of traditional staples, needed to acquire this language and technology to engage in the trade until such time as traditional sources of income were again available (Gilson 1970:254–259). The large numbers of Samoans who became involved in the production of cotton for a number of years, ensured that the vocabularies of cotton growing, harvesting and processing, plantation production, and marketing found their way into common use. Not all new crops, however, became central to commerce and generated expansion of related vocabularies connected with harvesting, processing and marketing. Although many things such as cardamom and cinnamon required new terms (katamome and kinamoni respectively), others such as the citron, jackfruit and peppers, could be rendered by combinations of existing Samoan words (moli patupatu, ulu suamalie and polo feu respectively).

Alongside this settlement, the construction of new types of buildings, using new designs, new materials, new fastenings, with a range of new tools gave rise to the same
expansionary processes described above. In building, which involved both missionary and settler activity, this technical language in some cases adopted Samoan words which had similar meanings (timber/laupapa; plane/olo); in others, Samoan words were given broader or additional meanings (spokeshave/fisi; brace and bit/‘auvili; nail/fao) and, in others, non-Samoan words were simply borrowed (‘mitre’/maita; set square/sekuea) and became part of Samoan. The gendered nature of some language expansion became apparent in this process. As missionaries encouraged their congregations to create new churches in which to worship their new God, it was primarily the men who had been traditional builders (tufuga) and labourers, who coined and used the terminologies for new building materials and technologies.

The missionaries, traders, sailors, and their dependants brought with them a new range of illnesses, diseases, treatments, and implements. Samoans who had, until this time, relatively little contact with the world outside Samoa had not had much opportunity to develop immunity to these new illnesses and they suffered as an increasing number of vectors introduced a series of new illnesses into the Samoan population. Unlike the situation which had prevailed in agriculture, and which had offered Samoans the opportunity to engage or not with the new phenomenon, widespread illness forced Samoans to engage with this new phenomenon. This required new ways of naming, conceptualising and responding to these illnesses. These changing patterns of illness presented Samoans with three choices: to stay with Samoan diagnostic and management paradigms, to adopt wholesale those of the new settlers who had clearly brought the illnesses, or to graft new beliefs and terms on to their pre-existing model of illness (Macpherson 1984; Macpherson and Macpherson 1990).

Since the illnesses with which Samoans were familiar were still present, there was no reason to abandon the ‘traditional’ paradigm, and since the new illnesses were so pervasive that they threatened the viability of Samoan society, they could not be ignored. The result was a rather ingenious fusion of the two paradigms. Illnesses which had always been in Samoa, and with which the Samoans were familiar, became a set known as ma’i Samoa for which concepts and a language were readily available; those diseases which had recently arrived, and with which the Papatalagi visitors were familiar, became a second set, ma’i Papatalagi, for which a new set of concepts and terminology was required. Terms for the new illnesses and for their treatments were accommodated in various ways. Some new conditions such as cancer (kanesa) required new terms while others such as high and low blood pressure (toto maualuga and toto maualalo) could be created by extending the meaning of existing words,1 and still others such as tuberculosis (māmā papala) could be incorporated using existing words without extension of meaning. Thus, over time, a number of new concepts and terms have made their way into the language of illness and treatment (Macpherson 1984; Macpherson and Macpherson 1990).

Throughout the latter part of the 19th century, new peoples arrived, passed through and settled in Samoa. This extended the range of sources of ideas and language. Some of these people were from parts of Europe: France, Germany, Austria, Scandinavia, some from the United States, and elsewhere in Polynesia.2 By the beginning of the 20th century a

---

1 The extension comes from the use of words for ‘high’ and ‘low’ respectively, that is measures of ‘height’ to refer to indices of ‘pressure’.

2 Some larger planters began to hire labour from Niue and the Cook Islands during the 1860s when they were unable to secure sufficient labour from Samoans (Gilson 1970).
significant number of new people from parts of Asia\textsuperscript{3} and Melanesia\textsuperscript{4} and a few from elsewhere in Polynesia had settled in Samoa. By 1914 there were some 3300 indentured labourers from these two sources in Samoa (O’Connor 1968). Despite attempts to separate these populations from the Samoans (Shankman 2001) a significant amount of contact occurred and with it came a new set of concepts and terms connected with everything from food to sorcery. A set of new ideas and, in some cases, the words for these moved into Samoan during this period and some, such as siaumeni (chow mein) and sopasui (chop sui) are now widely thought to be Samoan terms.

But this was not the only source of augmentation. Significant numbers of Samoan pastors and teachers, trained at the London Missionary Society’s theological college at Malua in Samoa, served with distinction in mission fields throughout Melanesia after 1846 (Gunson 1978). They had returned with new concepts, technologies and plants which they had found useful for medicinal and other productive purposes and which passed almost unmarked into Samoan society.\textsuperscript{5} During our study of Samoan medicine (Macpherson and Macpherson 1991), we would come upon small gardens of clearly exotic plants which people knew simply as plants from New Guinea or from the Solomon Islands but which had once been individually named.\textsuperscript{6}

Indeed, by this time, Samoans found it necessary to distinguish between the various sources of novel ideas and started to attach a series of qualifiers to mark the source. Thus, new varieties of banana which the missionaries brought from Kew Gardens (Musa balbisiana) became known as fa’i papalagi or ‘European’ banana (Parham 1972), while another from Niue (Musa paradisiaca) became known as fa’i niue; a variety of low fruiting sago palm from Rotuma (Metroxylon warburgii), whose leaves make excellent thatch, became known as niu olotuma (Whistler 2000); new edible nuts such as the cashew (Anarcardium occidentale) which arrived via India became known as the apu initia (Parham 1972). Flowering plants such as Plumeria obtusa from Hawai’i became known as pua hawaii while another cultivar from Fiji, Plumeria rubra, became known as pua fiti; and varieties of hibiscus from Tonga and Fiji were named, ‘aute toga (Hibiscus abelmoschus) and ‘aute fiti (Abelmoschus rugosus) to distinguish their origins (Whistler 2000). These distinctions were not reserved for valuable additions to Samoan flora: nuisances such as the creeper, fue saina (Micania micrantha), and a spreading rose, losa onolulu (Clerodendrum fragrans), both of which invade plantations; ringworm which came from the Tokelau Islands became lafa to’elau; and invasive creepers brought to Samoa on military equipment became known as vao malini or marine weed. But not all such adoptions are so marked.

---

\textsuperscript{3} Primarily from China, as plantation owners unable to secure stable supplies of Samoan labour began to recruit in China for Samoan plantations.

\textsuperscript{4} These labourers came mainly from the Solomon Islands and were engaged on large commercial plantations (O’Connor 1968).

\textsuperscript{5} Among these were plants such as ‘ava niukini (Derris malaccensis) which have become part of Samoan flora and fishing technology and might be assumed to have been endemic but for the qualifier niukini which marks its origins.

\textsuperscript{6} This tradition continues today as Samoan missionaries and others continue to return to Samoa with plants from mission fields in which they have served. We found for instance plants which had been introduced recently by Mormons who had served in reserve-based native American populations, and riggers who had worked with highlanders in remote locations in Papua New Guinea, in use in Samoan medicine.
Military activity was another source of new concepts and terminologies connected with new weapons, technologies and forms of social stratification. From 1902, when the Department of the Navy assumed responsibility for the administration of the eastern islands of the archipelago, Samoans were hired, first as labourers and later as tradesmen, in the naval base and shipyard in Pagopago. Early employees recruited other family members into the naval workforce until significant numbers of Samoan men were involved as civilian employees and later as service personnel. The military provided sets of terminologies, some relating to heavy engineering, welding, boiler-making, and shipbuilding in steel and iron; others relating to steam and later petrol and diesel propulsion plants; and others relating to rank and command structures. Service also produced a new means of ranking people: a new set of ‘titles’ which signified social rank and represented a new means of stratifying the world which bore no connection with genealogy or place: kapiteni/captain; satini/sergeant; sinia/senior.

Knowledge of these new arcane technologies and stratification systems conferred a certain social status on those men, who derived respect for their supposed knowledge of new ‘Papālagi’ technologies and the science which underpinned them. This was again gendered: new terms were more familiar to men who worked in the military, and to their families, who lived on or around the base and were not widespread. However, during the Second World War, which saw a large number of US soldiers with a new currency and new technologies stationed in Samoa for several years, a wider range of Samoans now came into contact with service personnel in Western Samoa as labourers, cooks, waitresses, laundresses and prostitutes. A number of adoptions and adaptations entered the language of many Samoans in a short time.

Military service also provided the first major opportunity for emigration from Samoa, and led to a process which was to have a profound effect on the Samoan language. When the US military withdrew from Pagopago to its larger bases in Hawaii in 1952, it offered Samoan military personnel the opportunity to relocate with their families, and some 2500 American Samoans migrated to Hawaii. Others later followed the first migrants and created significant migrant Samoan enclaves first around naval bases at Pearl Harbour and Honolulu, and later at bases in San Diego and San Francisco (Forster 1954). Migration, which began with the naval withdrawal, has since had a profound impact on the use of the Samoan language. The latter part of this paper considers the consequences of migration on the form of the language.

While the early impetus for Samoan migration was driven by US military decisions, a second set, connected with economic restructuring in New Zealand, soon led to a rapidly growing exodus to New Zealand and later Australia. There are now significant Samoan enclaves in cities in New Zealand (Pitt and Macpherson 1974; Fairbairn-Dunlop and Makisi 2003), Australia (Va’a 1995), Fiji (Tuimalealı’ifano 1990), Hawaii (Franco 1985; 1987) and California (Janes 1990). Migration of increasing numbers of Samoans to metropolitan nations around the Pacific Rim has produced a growing range of sources of language innovation. The second part of this paper deals with the ways in which these movements have impacted on the form and use of the Samoan language.

---

7 Before that Samoan people had left in small numbers as spouses, missionaries, trainees, and slightly larger numbers as people gathered around the temples of the Church of Jesus Christ of Latter Day Saints in both Lāie and Utah (Stanton 1978, 1992).
Some thoughts on change in the Samoan language

Diaspora and language shift

In these diasporic communities, the Samoan language has faced a range of challenges. Firstly, fewer and fewer people use Samoan language as a means of communication in significant parts of their daily life. This is not a matter of choice but of necessity. Despite legal protection for the use of Samoan at work, practical realities mean that English has become the language of the workplace for many men and women. Initially this trend was constrained because early migrants chose to work with other Samoans in places where the language of the workplace was Samoan, and to spend significant parts of their non-work life with other Samoans within the residential enclaves which formed around many cities and in which the Samoan language was used routinely (Macpherson 1974, 2002; Pitt and Macpherson 1974). In many respects, life in those enclaves during the 1950s and 1960s was very much like life in Samoan villages and the full form of the language was used, and heard, in a range of situations from informal discussion of food preparation to the formal oratory in the celebration of life crises.

The combined effects of a series of sociological factors meant that this situation could not last. The New Zealand case provides an insight into these factors. These included advice given, in the 1950s and 1960s, by influential ministers in Pacific churches to Samoan parents to speak to their children in English to ‘prepare them for life in New Zealand’; the cultural aspirations of some early Samoan migrants who sought to free themselves from the obligations of Samoan culture (Macpherson 1984, 1991); high levels of Samoan ethnic intermarriage (Macpherson 1972; Trlin 1975) which produced homes in which Samoan was only one of two or more languages used; the increasingly multicultural personal social networks of New Zealand-born Samoans in which English is used (Maingay 1995), and most significantly, the impact of English language medium formal education which effectively assigns a higher utility value to English.

The effect of these factors is reflected in the declining number of Samoans who are able communicate in Samoan. This is particularly so among the overseas-born descendants of the migrants, who represent steadily increasing proportions of the enclave population. In New Zealand, where language capability statistics are collected every five years, declining numbers of Samoans speak Samoan (Hunkin-Tuiletufuga 2001). This was evident in both the Samoan-born population, in which 90% could conduct an ‘every day conversation’ in Samoan, and in the New Zealand-born population in which only 46% claimed this ability (Hunkin-Tuiletufuga 2001). But this is not so much a matter of willful neglect, for Samoan migrants value their language and believe that it is the core of their culture (Hunkin-Tuiletufuga 2001), but rather of the restricted opportunities to hear, and use, the language in significant spheres of daily lives.

Among the first generation of New Zealand-born Samoans, inability to speak Samoan became the norm, and its cultural significance was readjusted. Samoan anthropologist, Melani Anae, explained that,

Despite their inability to speak Samoan fluently, most could understand Samoan. Their attitude was that language fluency wasn’t a ‘big deal’ because they felt they could pick it up at any time. Their ability to ‘tautala [speak] New Zealand-born’ enabled them to communicate with Samoan parents and island-born aiga, so there was no perceived urgent need to speak Samoan. They felt that because they had been raised by aiga and church in a Samoan way, had adopted parental fa’asamoa beliefs, values and attitudes and ways of thinking about and doing things, fluency in Samoan was not the single most important marker of Samoanness. (Anae 2001:110)
This was exacerbated by the growing trend of Samoan artists, musicians and writers to produce and perform material in English. Much comedy, popular music and rap is now performed in English by Samoan artists and musicians. Relatively few artists perform solely in Samoan, preferring instead to include one of two Samoan standards in their repertoire. This trend may once have been the consequence of decisions taken by marketing managers in Palāgi recording companies, but even Pacific-owned production companies such as Dawn Raid Entertainment, which ironically were established to promote Pacific artists, are now producing mainly English language rap and popular music for a steadily growing and increasingly prosperous New Zealand-born audience. Furthermore, as these Samoan artists are increasingly successful in national and international music industry awards, the English language formula becomes increasingly tempting to aspiring Samoan musicians.

The same is true also of Samoan writing, drama and comedy. While there is an increasing amount of Samoan writing, drama and comedy on offer in main population centres, which tell Samoan stories and address issues, there is little Samoan language writing, drama or comedy. As Samoan writers such as Albert Wendt and Sia Figiel, and film makers such as Sima Urale, David Fane and Oscar Kightley, and comedians such as the Naked Samoans, achieve national and international acclaim for their novels and poetry, films, plays and comedy respectively, the formula becomes attractive to Samoan writers and performers aspiring to reach a larger audience. As declining fluency becomes the norm, particularly among prominent and successful New Zealand-born Samoans, the incentive to speak declines.

But these trends may, ironically, be the consequence of earlier actions by Samoan speakers who demanded a high standard of Samoan language use and were openly critical of those who could not speak it well. These people frequently criticised and mocked those who called themselves Samoans but who used the language imperfectly. Young people who recalled these, often public, humiliations long after the event became increasingly wary about using the language in public. Often these events revolved around minor grammatical errors and mispronunciation which were repeated, creating a lot of fun for everyone except the unwitting victim. Some of these events and their consequences are recalled graphically in the accounts of adults whose families settled early in New Zealand (Fairbairn-Dunlop and Makisi 2003). Some noted that they would like to use the language more but that those early incidents robbed them of their confidence. The supreme irony is that the adult migrants who used these strategies to encourage their children to use the language well, in the process, have discouraged them from using it at all.

The symbolic significance of language as an ‘essential’ element of Samoan identity continues to decline for the reasons outlined above, and because New Zealand-borns have created an identity which has simply de-emphasised the importance of the language (Anae 2001). Many young Samoans choose to identify as Samoans and to include some Samoan language as elements of their Samoan-ness. But a relatively restricted set of words and phrases can suffice to indicate their symbolic attachment to both the language and culture of their parents. An emerging patois, the basic grammatical structure of which is English into which Samoan words and phrases are inserted, serves to identify one as a Samoan to both in-group and out-group. This style is increasingly popular and is characterised in the comedy of New Zealand-born Samoan comedians. It has another advantage, which some of its users noted in passing, which is that it is not immediately obvious to either Samoan or Papālagi strangers how much more Samoan one might know!
In other cases, New Zealand-born Samoans have opted for a ‘PI’ or ‘Poly’ identity which connects them with a larger group of descendants of Pacific migrants (Macpherson 1997). This choice effectively avoids challenges to the authenticity of their Samoan identities from Samoan speakers who discount their claims because of their inability to speak the language fluently. A PI identity cannot be challenged on these grounds and allows them to connect with other ‘Pacific People’ with whom they have much in common, including declining fluency in the languages, and familiarity with the cultures of their parents. Some connect with these groups precisely because their ability to speak these languages, or to know cultures intimately, is not central to the ‘validity’ of these PI identities.

A similar rationale is apparent in the increasing number of young Samoans who have transferred their religious allegiance from traditional, mainline Samoan churches to Pacific and general evangelical churches in which the language of worship is English (Macpherson and Macpherson 2001). Those who move do so for various reasons including the belief that they will not be able to take a role in church governance and leadership without Samoan fluency, and that they will always face challenges to their claims to be ‘real’ Samoans in these settings. Young people are also choosing to withdraw from formal association with churches as can be seen in steadily declining figures for religious affiliation in the census (Macpherson 2004). Since the Samoan church was, traditionally, the site in which many heard and learned certain formal elements of the Samoan language, throughout childhood and adolescence, the declining exposure to these settings in adolescence may well further limit the range of Samoan these young people hear and can use.

But not all have passively worked around these challenges to their fluency in Samoan. Some have responded to these challenges by parodying the English spoken by Samoan speakers. These parodies are found increasingly in the performances of New Zealand-born comedians, such as the acclaimed ‘Naked Samoans’ in both their stage shows and television programs such as bro’Town, drama, and in chatrooms where overseas-born Samoans meet. These parodies focus on the use of English by those who criticise their Samoan, and effectively shift the emphasis from the issue of their own fluency. The passage below, excerpted from a monologue circulating among New Zealand-born Samoans on the internet, counters criticism of their incomplete grasp of Samoan by pointing to their critics’ incomplete grasp of English.

Last week I was walking on the road. Wow! I saw a beautiful girl with a blue Samoan formal tunic-style dress. My feelings were carnal and I was dribbling at the mouth and wetting my own pareu. The scent of her perfume was so alluring. Her tunic framed her tummy. The way she walked brought me to want her desperately and I wanted her to become my chieftainess. Her figure was great, her back was fine. Her tunic set off her beauty; her face was lovely as was the red hibiscus flower that she wore. I wanted her to be my wife so I followed behind her. I know this is the first time [I have been in love] because I could not stop dribbling at the mouth.
Samoans in New Zealand are aware of and concerned by the declining use of the language. Organisations such as the association of teachers of Samoan, or Fāgāsā, have, with the support of the government and Samoan community leaders, taken measures to halt and reverse the trend. A number of radio stations in main centres broadcast a full range of programming in Samoan, and community access television produces some, mainly educational, Samoan language television programming. A nationwide FM network, funded by government and administered by the Ministry of Pacific Island Affairs, also broadcasts a range of material in Samoan.

All of these must, however, compete with commercial radio and television broadcasting which is more readily accessible to the growing numbers of young people whose ability in Samoan is restricted and who are interested primarily in youth culture. In an attempt to reach and to extend Samoan language audiences in their daily lives, commentaries of various sports are now broadcast in Samoan. Samoan language newspapers come and go, as have enterprises which have set out to publish Samoan language material for use in language education. The latter vary from comprehensive, multimedia language teaching programmes (Hunkin 1992) to colourful, pre- and primary school level readers.

Early childhood centres, ā‘oga ā‘amata, modelled on the Maori total immersion language nests, or kohanga reo, have been established and cater to a number of parents who seek to ensure that their children are fluent in Samoan. These centres are particularly valuable to parents who do not have Samoan and cannot themselves teach their children at home. But these centres can cater only for children in areas where there are populations large enough to support them, and compete with a lingering belief among some parents that English fluency is the key to upward social mobility. At present, and despite growth in the number of enrolments, only about 25% of Samoan children in early childhood programmes are enrolled in ā‘oga ā‘amata (Hunkin-Tuiletufuga 2001:206–207). There have been a series of initiatives to teach Samoan in schools, universities and community education programmes which have been used by Samoans who did not learn the language as children, and non-Samoans who are married to, manage, or work with Samoans and seek to learn the language to better understand the culture. But despite moves designed to reverse the decline in use (Hunkin-Tuiletufuga 2001:210–211), this trend seems set to continue for a range of symbolic and practical reasons outlined above.

At one time, the declining use of Samoan looked likely to be a phenomenon of communities of the diaspora. But return migration, for shorter and longer terms, has ensured that linguistic innovation has found its way back into Samoa. This process has been occurring for some time (Holmes and Holmes 1992) but its speed has increased as new telecommunications and media have made it possible to transmit huge volumes of material to Samoa within seconds of its production anywhere in the world, and to reproduce it in Samoa at a very low cost. Combined with the greater numbers of returning migrant Samoans and more and more media, this process gains momentum by the year. Now from a sitting room in Savaii one can watch Television One News from New Zealand every night of the week; the popular soap opera Shortland Street five nights a week, and the Pacific magazine programme ‘Tangata Pasifika’ weekly. Between these one can watch documentaries from BBC, ABC, and CNN and films from all over the world, including the increasingly popular Bollywood epics, on Samoa’s free-to-air national television network. The one thing that these have in common is that except for locally produced news, educational and some entertainment programs, few are in Samoan. When one tires of that
diet, DVD movies from the US, New Zealand, Australia, Britain and India can be hired from village video stores throughout Samoa.

This increased exposure may lie behind the fact that ideas and concepts which generally can be (and have been) expressed adequately using combinations of words which exist in Samoan are increasingly rendered as a phonologically adapted version of the original term or indeed as the original term. Thus a taxi, which was once rendered *se ta’avale la’u pase se* ‘a car carrying paying passengers’, is increasingly referred to as simply *se taxi*. This process is increasingly rapid. Cellular phones which became available only two years ago were referred to originally as *telefoni feavea’i* ‘phones which could be carried around’, only to become *selulā* and more recently *seli*. On the surface this seems to be in the interests of ‘linguistic economy’, but the net effect of this phenomenon may be the loss of the ability to render these ideas in Samoan.

The number of terms which fall victim to this process of creep seems to increase for a number of sociological reasons. New non-Samoan terms and expressions, which were once parts of esoteric technical vocabularies, and were used only by a limited number of people in professional discussions are increasingly finding their way into a more popular discourse. This occurs as more people are drawn into new spheres in which the use of these new terms is routine. In agriculture, for instance, phonologically adapted Samoan and English terms were once used only by agricultural advisors and large-scale farmers who could afford to use expensive herbicides. Most small farmers referred to the sprays as *vai lā’au* ‘medicine’. As small farmers become increasingly engaged in commercial agriculture, the differences between various herbicides become critical to profitability and now, in and around farm stores, even small farmers are engaged in earnest conversations over the relative merits of ‘Sting’, ‘Roundup’ and ‘Glyphosphate’ and the benefits of applying the sprays with ‘wands’, ‘backpack pump sprays’, or ‘motorised backpack sprays’.

Part of the shift is a consequence of the linguistic conduct of Samoan opinion leaders. Members of the social and political elites, even in Parliament, increasingly use these abbreviated and adapted words when commenting on policy and current events in the mass media in ‘the interests of communication’ and on the not unreasonable assumption that their listeners will understand them. In some cases, they go further than simply using adapted versions and add, …..ua ta’ua 1 le fa’apalāgi ole differential taxation ‘which is called in English differential taxation’. While on one hand, this tendency achieves clarity, it also implies that Samoan, even in the hands of the elite, cannot effectively express an idea such as differential taxation in Samoan.

Part of the shift is a consequence of Samoans’ passion for sport. Radio and television broadcasting in the Samoan language has produced an explosion of new terms in sports like rugby, netball and horse racing. The commentators justify the use of the large numbers of adapted terms in these commentaries as essential since the action occurs so quickly that to describe the action in Samoan would be impossible, because it would take too long and would annoy some listeners. Others argue that since Samoans can watch so much televised sport in English they are familiar with and used to non-Samoan terms and have no problem with the hybrid commentary.

A generation of Samoan children growing up in Samoa with these terms is now unaware of the earlier forms of expression. A taxi driver asks me in Samoan if I would like to listen to the radio (*leitio*), a tape (*se lipini*) or a CD (*siti*). A colleague teaching at the National University of Samoa reports that an increasing number of Samoan students are
asking to enrol in Samoan language courses offered at the university as a foreign language which suggests that the process may have moved more quickly than is assumed.

Does this all matter? Probably not. As I noted at the outset, the Samoan language has been changing since its inception, and one might argue that this is simply part of an inevitable process of social and cultural change which will occur simply because Samoa is increasingly influenced by homogenising global forces. The pragmatist might argue that Samoans seem to manage perfectly well with the pace and direction of change in their language. Even Samoa’s most important scholar of Samoan, Tuiatua Tupua Tamasese Efi, acknowledges that languages change,

> We constantly need to search for meaning, nuance and metaphor to find substance and establish context in our dialogue with our ancestors, with ourselves and with other cultures. In this process, we need to recognise that change and pluralism are part of life. A living culture cannot be sustained by ritual or measure that is divorced from the modern contemporary context. (Efi 2003:24)

This may be true as long as the additions and transformations of Samoan add to rather than displace concepts. In another paper, on the future of Samoan in the 21st century, Tuiatua Tupua Tamasese Efi identifies four measures that are necessary to ensure that the language survives in a form which allows for dialogue with the ancestors (Efi 1995:2). It is salutary to reflect that eleven years after this paper was delivered to an audience in Samoa which might have been expected to embrace the significance of these measures, very limited progress has been made on any of these, and that the second of these, the teaching of Samoan in pastors’ and catechists’ schools has effectively ceased, at a time when the language is under more pressure than ever before. It may be that only linguists will ultimately worry about these issues and the process of which they are a part.

References


—— 1974, Toward an explanation of the persistence of extended kinship among Samoan migrants in urban New Zealand. DPhil thesis, University of Waikato.


Parham, B.E.V., 1972, *Plants of Samoa. A guide to their local and scientific names with authorities; with notes on their uses, domestic, traditional and economic.* Wellington: New Zealand Department of Scientific and Industrial Research.


Powell, Thomas, 1886, *O le Tala i Tino o Tagata ma Mea Ola Ese'ese e iai fo'i o tala i manu ua ta'ua i le Tusi Pa'ia.* Chilworth: Gresham Press.


Proto Polynesian *kainaŋa
‘1. the populace of commoners;
2. matrilineal descent group’

JEFF MARCK

1 Introduction

This paper argues that Proto Polynesian and Proto Micronesian had the form *kainaŋa which meant ‘matrilineal descent group’ but was, in Proto Polynesian, polysemous and had a more general meaning around Tonga and Samoa of ‘the general populace: commoners as opposed to people from chiefly families (*qariki)’. The paper asserts that ancestral Polynesian society (=Proto Polynesian society) had no single system of descent practised by all elements of its population. The general populace is modelled as having emerged into the cognatic (bilateral, non-unilineal, ambilineal) descent system that typifies most Polynesian societies today. But some elements of the population, probably those amongst whom seafaring was still most important, were still practising the matrilineality of their Proto Oceanic society forebears and were perhaps still matrilocal as well. The consideration of Polynesian and Micronesian *kainaŋa forms was becoming a major digression in another work (Marck 2008) and I was happy to think I could offer this study separately in celebration of Andrew Pawley’s career.

Map 1 shows Polynesia in the Pacific and the distribution of the Polynesian subgroups. Figure 1 gives the subgroups of Polynesian. The Polynesian Outliers are inhabited by Polynesian speaking groups whose linguistic forebears arrived anciently from within Triangle Polynesia to localities in Melanesia and Micronesia stretching from West Uvea in the Loyalties (north-east of New Caledonia) through various Vanuatu and Solomon Island localities up to Nukuoro and Kapingamarangi south-west of Pohnpei in Micronesia. Along with Pukapuka, where there is a non-East Polynesian language in western East Polynesia, two of the Outliers will be of special interest here: Anuta and Tikopia which are in the south-easternmost area of the modern nation of the Solomon Islands. The Outliers from Anuta and Tikopia south comprise the Futunic Outlier group, named due to their apparent origin in East Futuna (Bayard 1966, 1976; Pawley 1967) or perhaps East Uvea or both.

1 My work on this paper from 2006 to 2008 was supported by Australian Research Council Grant DP0663170 ‘Proto Oceanic language, people and society: foundations of the Austronesian settlement of the Pacific’, directed by Andrew Pawley and Malcolm Ross.
The Outliers north-west of Anuta and Tikopia are known as the Ellicean Outliers (cf. Howard 1981), so-named due to the apparent origin of their speech in Tuvalu (called the Ellice Islands until independence).

<table>
<thead>
<tr>
<th>Tongic</th>
<th>Nuclear Polynesian</th>
<th>Tuvaluan and Ellicean Outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tongan</td>
<td>Pukapukan</td>
<td>Tuvaluan</td>
</tr>
<tr>
<td>Niuean</td>
<td>Futunic</td>
<td>Nukuoro</td>
</tr>
<tr>
<td>Nuclear Polynesian</td>
<td>East Futunan</td>
<td>Kapingamarangi</td>
</tr>
<tr>
<td>Pukapukan</td>
<td>Anutan</td>
<td>Takuu</td>
</tr>
<tr>
<td>Futunic</td>
<td>Tikopian</td>
<td>Nukumanu</td>
</tr>
<tr>
<td></td>
<td>Pileni</td>
<td>Nuguria (Nukuria)</td>
</tr>
<tr>
<td></td>
<td>Rennell – Bellona</td>
<td>Luangiuia (Ontong Java)</td>
</tr>
<tr>
<td></td>
<td>Mae</td>
<td>Sikaiana</td>
</tr>
<tr>
<td></td>
<td>Mele-Fila</td>
<td>Samoan and Tokelaun</td>
</tr>
<tr>
<td></td>
<td>West Futuna – Aniwa</td>
<td>East Polynesian</td>
</tr>
<tr>
<td></td>
<td>West Uvean</td>
<td>Rapanui (Easter Island)</td>
</tr>
<tr>
<td>Ellicean</td>
<td>East Uvean</td>
<td>Central East Polynesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marquesic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hawaiian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nuclear Marquesic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marquesan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mangarevan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tahitic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tahitian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NZ Maori</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cook Island Maori</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tuamotuan dialect region</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Austral dialect region</td>
</tr>
</tbody>
</table>

**Figure 1:** Subgroups of Polynesian
Figure 1 shows that Polynesian languages fall into two primary groups: Tongic and Nuclear Polynesian. Proto Nuclear Polynesian disintegrated with the differentiation of Pukapukan (a non-East Polynesian language in western East Polynesia), East Uvean, East Futunan and thence the Futunic Outliers, and Ellicean (a group composed of Samoan, Tuvaluan, Tokelauan, the Ellicean Outliers and East Polynesian). The splits that will ultimately most concern us below involve the era of initial Nuclear Polynesian disintegration and the status of Pukapukan and two Futunic Outliers (Anutan and Tikopian) as they are the only Polynesian languages which retain a ‘descent group’ sense in their *kainaŋ* forms. The splits are presently dated to the early first millennium AD in a process which continued the disintegration of a unified language around Western Polynesia from the Proto Polynesian period (Pawley 1996).

Map 1: Polynesia in the Pacific and its linguistic subgroups

2 The Proto Oceanic social structure heritage

Proto Oceanic society is defined as the society of the speakers of Proto Oceanic, who inhabited a network of interacting communities situated around the Bismarck Archipelago at about 1200 or 1000 BC. Proto Oceanic was ancestral to all Austronesian languages in the Pacific Islands except Chamorro and Palauan of Western Micronesia. Proto Polynesian society and Proto (Nuclear) Micronesian society were descended and transformed societies that existed, circa AD 1 around Western Polynesia (Tonga, Samoa and the smaller islands
around them) and Central Micronesia (Chuuk, Pohnpei, Kosrae and the smaller islands around them) respectively.

The only recent model of Proto Oceanic descent systems began with Hage (1998a, 1999) who asked if Proto Oceanic society was matrilineal. His suggestion that it may have been matrilineal was based upon indices of matricentricity, lexical markedness and other factors. That suggestion soon intersected with studies of Polynesian’s human genetics which showed sex-biased gene flow typical of matrifocal residence from the period of their ancestors’ residence in Melanesia (Hage and Marck 2003; Kayser et al. 2006). Given Kayser et al.’s (2006) present interpretation of Polynesian mitochondrial and Y-chromosome DNA, I have suggested that average Polynesian ancestors’ unions with Melanesians per generation during their centuries in Melanesia were in the order of one in ten fathers per generation being indigenous Melanesian but less than one in one hundred mother per generation being indigenous Melanesian (Marck 2008). Men, Kayser et al.’s interpretation suggests, were marrying ‘in’ but women hardly did so at all, a pattern consistent with matrifocal rather than other forms of residence. This suggests that a child’s language/cultural learning through the Polynesian ancestors’ centuries in Melanesia occurred in a context where, on average, something like 95 percent of adults in their communities were native speakers of early Oceanic and likewise native culture bearers of early Oceanic society (Marck 2008).

As is typical of historical matrilocal societies (matrilocal residence with formal matrilines) and uxorilocal societies (matrilocal residence without formal matrilines), Kayser et al.’s (2006) study and earlier work by Kayser and others summarised by Hage and Marck (2003) suggest that in Proto Oceanic society there was simply little place for outside women while outside men presented no similar contradictions with respect to a ‘place’ for them in residence and inheritance practices. Outside men, like men from within the group, resided with their wives or in men’s houses and their children inherited their land rights from their mother rather than their father.

It is convenient to now mention a certain distinction in social organisation terminologies. This is the hard, fast and narrow definition of ‘clan’. While indigenous Polynesians sometimes refer to their own ‘clans’, Polynesian societies are mainly without clans in the social organisation sense. In the parlance of social organisation studies, ‘clan’ always refers to unilineal (matrilineal or patrilineal) groups where membership is simply, completely and universally defined in terms of either one’s maternal or paternal line.

---

2 Where Murdock’s (1949) earlier notions of a cognatic Proto Oceanic Society are dismissed.

3 The terms ‘virilocal’ and ‘uxorilocal’ were introduced by Adam (1948) as truer to the Latin roots than ‘patrilocal’ and ‘matrilocal’ since one is attempting to specify residence in reference to spouses rather than their children. Many dictionaries of English and even dictionaries of anthropology maintain that ‘patrilocal’ and ‘virilocal’ remain synonyms as do ‘matrilocal’ and ‘uxorilocal’. However, Murdock (1967) used ‘virilocal’ to tag societies which had patrifocal residence in the absence of patrilineages and ‘patrilocal’ to refer to societies which had patrifocal residence in the presence of patrilineages. Similarly, he used ‘uxorilocal’ to denote societies which had matrifocal residence but no matrilineages and ‘matrilocal’ to denote societies which had matrifocal residence as well as matrilineages. If Murdock’s definitions are not now universally employed they are convenient and the senses in which I use the aforementioned terms in the present work are as Murdock defined them in 1967.

4 See Sahlins (1958:139–141) for a consideration of how the ‘clan’ term had been used up until that time in Polynesianists’ literatures.
Polynesian societies rarely calculate membership in their kin groups by unilineal criteria and most are classed as cognatic.

Firth (1957) has considered how the cognatic systems of Triangle Polynesia are sometimes profoundly patrifocal but patrilineality is rarely achieved since exclusive lines of men are never formalised into patrilines (the exceptions are Anuta, Tikopia, Rennell/Bellona and Pukapuka). As Firth explained and generations of Polynesianists have observed, the occasional daughter and her husband in even highly patrifocal Polynesian groups may raise their children on the land of that woman’s father or mother’s group, the children emerging as normal/ordinary members of that land and their (non-unilineal) descent group depending on activation, by virtue of residence, of those options as opposed to alternate residence and cognatic descent group options.

Of the four Polynesian societies which are formally lineal, three, Anuta, Tikopia and Rennell/Bellona, are Futunic Outliers and all three are patrilineal. Anuta and Tikopia have the *kainaŋa form and use it to refer to their patrilineal clans. Pukapuka, the fourth, has ‘double descent’ (both matrilineages and patrilineages) and one level of matrilineal sub-lineage is the keinanga, a matter that shall be reviewed presently in light of the agreement of that form and meaning with Micronesian *kainaŋa ‘matrilineal descent group’ forms. Double descent systems are one form of transition from matrilineality to patrilineality or vice versa and are rather rare amongst Oceanic speaking groups (Marck 2008). I view the Pukapuka case as an example of matrilineal to patrilineal transition (or stasis). Pingelap in Micronesia is another double descent case, one which has clearly arisen out of a matrilineal past. The Pingelap *kainana form specifies the patrilineages rather than the matrilineages, as would seem to have happened in the past of Anuta and Tikopia, *kinaŋa probably referring to both at some transitional stage.

Having considered how rare lineality is in Polynesian societies, I return to the subject of the Proto Oceanic descent situation. All subgroups of Oceanic other than Southern Oceanic (Southern Vanuatu and New Caledonia) are known to have daughters whose speakers’ societies include matrilineal groups (Marck 2008). If matrilineality appears absent in Southern Oceanic, we must allow that such may simply be due to lack of description. Everywhere matrilineal societies existed upon earliest historical description amongst Oceanic speakers, their distribution seems be retreating or is in stasis rather than advancing.

This distributional situation is very strong evidence in favour of the notion that Proto Oceanic society was matrilineal/matriloclal rather than cognatic/uxorilocal (Marck 2008). The distributional situation was mentioned in passing but not emphasised by Hage (1998a) and is the subject of special attention in Marck (2008).

Change in descent commonly follows changes in residence amongst human societies (Murdock 1949). Shifts in residence to matrifocal systems are commonly the result of the regular absence of men (seafaring, warrior societies, traders, herders) (Hage and Marck 2003:S123) and matrifocal residence is also a common adjustment to migration to new lands and conflict with the ‘others’ there encountered (Divale 1984). As the population was both seafaring and migratory, Proto Oceanic society was composed of people who had two motives for matrifocal residence. Matrifocal residence rarely persists without the eventual development of formal matrilines (see Divale (1984) and also observe the general rarity (Murdock 1967) of uxorilocal systems in human societies compared to matrilocal

---

5 Kirch and Green (2001:212) give Bellona kakai ‘anga in their table of *kainanga forms but I do not consider it cognate.
societies). The demise of matrilocality amongst Oceanic speaking societies is associated with abandonment of long-distance seafaring amongst local populations in Micronesia (Hage and Marck 2002), is universal in Polynesia⁶ and is common through Melanesia (Marck 2008). The distributional argument notes that matrilineality is found amongst at least some speakers of perhaps all the Oceanic subgroups. This distribution is predicted to be retreating by the general model due to the decline of its original sources: matrilocality due to seafaring and migration. And matrilocality and matrilineality are, in fact, observed to be retreating nearly everywhere they are found (Marck 2008).

Although all subgroups of Proto Oceanic other than Southern Oceanic are known to have daughters whose societies include matrilineal groups, the name for ‘matrilineal descent group’ or even just ‘clan’ is shared between no two Oceanic subgroups⁷ except in the case of Micronesian and Polynesian, which share *kainaŋa.⁸ The view taken here is that the Micronesian forms always meant ‘matrilineal descent group’ as does the Pukapukan form in Polynesia (where, specifically, it names matrilineal sub-lineages in the double descent system).

Although the Pohnpeic Micronesian forms have a shared irregularity (see below), there is no apparent source for the irregularity through borrowing and no borrowing hypothesis has ever been published. From a social organisation perspective one might wonder why one need be looking for a borrowing explanation in any event. The ancestral society was matrilineal and matrilocal at the level of Proto Oceanic and at any later common interstage ancestral to both Polynesian and Micronesian. Why would it be borrowed from Micronesia to Polynesia where lineality waned from such an early time? Why would it be borrowed from Polynesia to Micronesia when the source populations were shedding lineality and the target populations already had vigorous forms of matrilineality?

### 3 History of ideas concerning *kainaŋa

The earliest source to speak to the ancient meanings of *kainaŋa seems to have been Goodenough (1955) who conflated reflexes of Proto Polynesian *kainaŋa and *kaaiŋa, the former having to do with people in Proto Polynesian, the latter having to do with place. I have argued that the former has cognates in Micronesia due to common inheritance from early Oceanic while the latter has no apparent cognates outside Polynesian, Kiribati kainga, the only presently identified candidate, being best explained as a Polynesian loan (Marck 1996:205, 2000:188–189). Goodenough mentioned Chuukic (Micronesian) cognates of Polynesian *kainaŋa and considered them variants of *kaaiŋa ~ *kainaŋa.

The next mention of *kainaŋa in the literature seems to be Koskinen (1960:157–158) where the Polynesian forms *kaaiŋa and *kainaŋa were seen as distinct, the latter having to do with ‘populace of a place’ and the former possibly originating in ‘the verb kai ‘to eat’’. Koskinen noted that the *kaaiŋa forms sometimes referred to just place and sometimes to both places and their people, a matter I will return to presently.

---

⁶ Pukapukan society is patrilocal (Beaglehole and Beaglehole 1938:230).
⁷ But see Chowning (1991:69) for some faint evidence of an early Oceanic *qapusa ‘clan’.
⁸ Another kith and kin term has a similar distribution, is also ancient, perhaps to the level of Proto Oceanic itself, and also involves an agreement limited to Micronesia and Fiji/Polynesia but for the exception of a single known Melanesian cognate from Buka immediately north of Bougainville (Blackwood 1935:56): pasu ‘man’s sister’s child’ (Proto Micronesian *fasu ‘man’s sister’s child’, Proto Fijian *vasu ‘man’s sister’s child’, Proto Polynesian *fasu ‘privileges of a man’s sister’s child’).
Koskinen (1967:95) again discussed *kainaŋa forms and may have erred in terminology when he stated that in ancient Polynesian society, ‘[l]ineage was the basic structure of the community.’ Like ‘clan’, ‘lineage’ is a term reserved in social organisation usages to situations where there is unilineality (patrilineality, matrilineality or double descent) and it seems unlikely that Koskinen would have meant to imply that of the mainly cognatic Polynesians. More probably Koskinen used the term in the looser sense some Polynesianists occasionally have, subsuming cognatic systems for which there is no common equivalent term to ‘lineage’. In any event, Pawley (1982:43) employed a more appropriate usage when he quoted Koskinen as I have above but in the next sentence spoke of Polynesian ‘descent’.

Goldman (1970:544) was aware the *kaaiŋa word first had to do with place but, like Koskinen, wondered if it also anciently could have had to do with one’s descent group. But the ‘descent group’ meaning is limited to Western Polynesia⁹ and is now interpreted by Biggs and Clark (2006) as having been borrowed through that area after the dispersals to the Outliers and East Polynesia. Goldman didn’t discuss *kainaŋa forms in his conclusion though he noted historic senses in some of his earlier chapters on individual societies. One of those chapters considers Pukapuka and mentions the matrilineal keinanga of Pukapukan society. Goldman (1970:392) puzzles that ‘[t]he keinanga are inexplicably exogamous, and have tended to take upon themselves the actual economic authority over the taro beds.’ Precisely the same is true of the *kainaŋa of the Chuukic speaking atolls of Micronesia (cf. Kuehling (n.d.) for Woleai). The model proposed here attributes those identities to common heritage from the early Oceanic era. It also attributes those identities to the societies of the interstages involved: Proto Micronesian, Proto Pohnpeic-Chuukic, Proto Pohnpeic and Proto Chukic in the Micronesian line and Proto Central Pacific, Proto Polynesian and Proto Nuclear Polynesian in the Polynesian line.

Pawley (1982:44–45) considers *kainana forms from Polynesia and Micronesia, suggesting there and again in Pawley (1985:96) that it was probably formed from the early Oceanic root *kai(n) ‘people of a place’. Verbal forms of the word apparently included *kai and *kakai ‘be inhabited, settled’ and *kainaŋa is seen to be formed of *kaina-, a verbal construction, and *-ŋa, a Proto Oceanic nominaliser. Pawley (1982:44) reconstructed Proto Polynesian *kainaŋa ‘land-holding descent group’ as well as ‘subjects of a chief, the common populace as opposed to those of noble birth’ and three years later (Pawley 1985:96) reconstructed Proto Polynesian *kainaŋa ‘lineage or clan, people acknowledging same ancestor and chief’. Comparing Micronesian and Polynesian data, Pawley (1985:96) concluded that an early Oceanic *kainaŋa ‘high order descent group’ should be reconstructed.

The next consideration of *kainaŋa’s Polynesian past appears to have been Biggs and Clark (1994) where a semantic reconstruction was made: ‘some social group, perhaps commoners as opposed to aristocracy’. Two years later I reconstructed Proto Polynesian meanings of ‘clan; worshippers of a deity; subjects of a chief; commoners’ (Marck 1996:204). I suggested (Marck 1996:204; 2000:186–187) that the various meanings developed over time:

These senses become comprehensible in the context of Polynesian social and religious practices whereby the most commonly worshipped deities were family ancestors, and the most common kind of chief was the ranking male of the kin group (however that group was constructed). Thus if PPn did have clans called *kainaŋa, it is not difficult to imagine that they were subject to the same chief, and

---

⁹ Except for the Futunic Outlier Pileni which has a ‘family’ sense in addition to its ‘village’ sense.
gathered together to worship the same deities under the leadership of that (clan) chief. The ‘commoner’ sense may have developed over time as the most senior lineages of the *kainaŋa came to be remote chiefly people and the old name for the group itself came to refer to the mass of its members who were of common rank.

That wording rescues me from some direct conflicts with my present opinion but only just. I now view the Proto Polynesian (unilineal) ‘clan’ sense as a retention most likely having to do with the populations of the smaller, more isolated settled islands around Western Polynesia10 (East Uvea and East Futuna, perhaps) and perhaps residually amongst specialist sea transport and fisher folk around Tonga and Samoa. This was occurring in a context which Pawley (1996) describes as Proto Polynesian’s drift into southern (Tonga) and northern (Samoa and the others) speech varieties around Western Polynesia rather than in sudden, definitive breaks in either linguistic or other cultural continuities. I did, however, err (Marck 1996:206–207; 2000:190) when I suggested that the social anthropologists would have to argue the nature of the Proto Polynesian clan and ‘[w]hether to reconstruct it as patrilineal, matrilineal, nonunilineal, ambilineal, or something else.’ As I came to understand from the time I began working with Hage in 1999, it was wrong to use ‘clan’ and ‘nonunilineal’ or ‘ambilineal’ in the same breath. And, as I now believe, the word had two distinct senses in Proto Polynesian: 1. the general populace: commoners as opposed to chiefly families (*qariki) and 2. matrilineal clan in a mixed descent system.

The Proto Oceanic matrilineality model now implies that by Proto Polynesian times the drift to cognatic systems through Western Polynesia may have been advanced. However, the evidence of Pukapukan, Anutan and Tikopian is clear: at least some elements of the Western Polynesian Proto Polynesian speaking populations continued to calculate unilineal descent and to use *kainaŋa to refer to clans, lineages or sub-lineages.

Hage (1998a, 1998b) published what may now seem contradictory opinions. Citing my invitation to the social anthropologists mentioned above (to consider the kind of clan Proto Polynesian *kainaŋa was) he put forward the opinion that it was patrilineal (Hage 1998b:189) based upon that sense being current in Anutan and Tikopian and due to the otherwise common patrifocal nature of descent groups amongst Polynesian societies. But after the publication of his Proto Oceanic society matrilineality hypothesis (Hage 1998a) and especially after the human genetic evidence elevated the hypothesis to a firmer model (Hage and Marck 2003), he (pers. comm.)11 came to the opinion that the earliest Polynesians were matrilineal and that most patrifocal cognatic Polynesian societies had become such in periods that emerged out of matrilineal rather than patrilineal ancestral societies.

Hage (1999:202–203) only touched upon existing ideas about *kainaŋa in the literature in his general overview of Proto Oceanic society. He took the position that Proto Oceanic society was lineal, either patrilineal or matrilineal, but did not press the (Hage 1998a) matrilineal case, mentioning it only in the final paragraph before the postscript, apparently due to publishing lag. Hage (1999) was developing the lineality hypothesis more than the matrilineality hypothesis as a lineality hypothesis was a necessary first step. Similarly, Hage and Marck (2002:151) only mention ‘Proto Oceanic’ *kainaŋa ‘descent group’ and then consider its best definition in Proto Micronesian (‘matrilineal descent group’) in that bottom up reconstruction of the Micronesian situation.

---

10 The general area of Proto Polynesian speech (Pawley 1996).
11 Per Hage died in 2004 (Milicic 2006; Marck 2006) just as many of his opinions were evolving.
As did Hage (1998b), Kirch and Green (2001:211–218) took up my 1996 challenge to form a social organisation opinion of *kainaŋa’s meaning in Proto Polynesian that went beyond my tentative opinions based upon linguistic methods alone. Kirch and Green (2001:213) failed to recognise their Rotuman form/evidence as a Polynesian loan, but that does not impact on their most basic conclusion: that Proto Polynesian *kainaŋa meant ‘land-holding or controlling group tracing ascent from a common ancestor’. ‘We also infer that these groups were exogamous, and that they were likely to have been unilineal, although we cannot say with certainty whether the principle of ‘ascent’ was matrilineal (as in Pukapuka) or patrilineal (as in Tikopia)’ (Kirch and Green 2001:214). As I do in the present report, Kirch and Green (2001:214) argue that ‘the term and its associated meanings underwent significant changes in both Western and Eastern Polynesia after the breakup of the PPn speech community.’ The difference here is that I allow that descent in Proto Polynesian society was not necessarily of a homogenous form and that many of the changes surrounded elements of the society which were no longer matrilineal and seem to have been cognatic by Proto Polynesian times.

4 Proto Polynesian *kainaŋa revisited

First we must dismiss any notions that people with a homogenous language must have a homogeneous descent system. We have no basis or method in linguistics to demand that such be the case and neither can we demand such from the perspective of synchronic studies of social organisation. Indeed, societies may have diverse descent systems along several axes. There are sometimes differences along the axis of rank as in the case of certain groups in the Admiralties:

Chinnery (1925:52–54) includes materials on Aua and Wuvulu (Durour and Matty islands), the “Western Islands” of the Admiralties. He describes a situation where the commoners are matrilineal and matrilocal but the chiefs more ambilateral and the highest chiefs patrilineal. (Marck 2008)

Similar to the Aua and Wuvulu differences along a rank axis, there are sometimes differences in South Asia along the axis of caste. The Nayar of Kerala, famously polyandrous,12 are matrilocal but part of a larger regional society where other groups are not.

Geographical variability in descent systems would be an obvious consequence of the vast geographical spread of Proto Polynesian speech (all of Western Polynesia settled at the time)13 and there seem to have been pan-Western Polynesian developments that affected the general populace in new ways while some smaller groups retained the old. I now posit geographical and perhaps occupational variation in Proto Polynesian society’s descent systems whereby East Uvea or East Futuna or both were still matrilineal, as were perhaps some seafarer specialists on Tonga and Samoa. By Proto Polynesian times *kainaŋa still meant ‘matrilineal clan’ to the populations that were still matrilineal but amongst the majority of Tongans and Samoans the word was no longer so clearly associated with descent and had come to designate the general population of commoners.

First there is the evidence that there was a Proto Polynesian sense of ‘matrilineal descent group’ retained from early Oceanic and that this meaning continued in at least the ancestor of Pukapukan and, perhaps from a different geographical area of Western Polynesia, the ancestor of Anutan and Tikopian. After almost ten years, Hage’s (1998a) Proto Oceanic

---

12 A woman may have multiple husbands at the same time.

13 It wasn’t, perhaps, until after Proto Polynesian times that Tuvalu and the Tokelaus came to be settled.
society matrilineality hypothesis has been elevated to a strong model due to genetic studies (Hage and Marck 2003; Kayser et al. 2006). It is also strengthened by the distributional evidence (Marck 2008) mentioned above, and is uncontested in the literature. Our task then becomes one of surveying the mainly cognatic landscape of Polynesian societies for any vestiges of matrilineality. In Pukapuca we find not only matrilineality but an early Oceanic word for ‘matrilineal descent group’ in Pukapukan keinanga ‘matrilineal sub-lineage’. We also find an exact match for two of the keinanga’s social functions (exogamy and ownership of the taro swamps) in Proto Chuukic *kainaŋa. So the presence of matrilineal descent groups amongst elements of the population is demonstrated for Proto Polynesian society. As previously mentioned, the ‘patrilineal’ senses of Anuta and Tikopia are then seen as the likely result of periods of double descent followed by the abandonment of matrilineal. I draw attention, again, to Micronesia’s Pingelap *kainaŋa form which has come to mean ‘patrilineal clan’ out of a more conspicuously matrilineal past than Polynesia’s Anuta and Tikopia.

I now turn to the second sense of Proto Polynesian *kainaŋa, that of ‘populace, commoners’ and related meanings seen in the literature reviewed above.

Linguistic evidence suggests reconstruction of Proto Polynesian, Proto Nuclear Polynesian, Proto Ellicean, Proto East Polynesian, Proto Marquesic and Proto Nuclear Marquesic *kainaŋa ‘commoner, general populace’:

- Tongan kainanga ‘populace, people without chiefly rank’
- East Uvean kainanga ‘people not of chiefly rank’
- Hawaiian maka-‘eina ‘commoner, populace, people in general’
- Marquesan mata-‘eina ‘people, gens, sujets’
- Mangarevan mata-kainanga ‘assembly, congregation of persons’

Linguistic evidence also allows a second reconstruction not clearly attributable to Proto Polynesian or Proto Tongic but attributable to Proto Nuclear Polynesian, Proto Futunic, Proto Ellicean, Proto East Polynesian and Proto Tahitic: *kainaŋa ‘attendants and subordinates’:

- Samoan ainaga ‘attendants and minister of the aitu (spirits, gods)’
- Mae (a Futunic Outlier in Central Vanuatu) na-kainanga ‘titled person subordinate to a given person’
- Tahitian eina ‘female attendants of the queen or chief woman’
- Tuamotuan keina’a ‘a group, band, body of followers, servants, people united by the same service, duties; the female attendants of a chieftainess’

Two distributional observations are notable. First, the living languages with the ‘commoner’ and ‘subordinates’ senses never use their *kainaŋa forms to name any sort or level of descent group and the living languages that have a ‘descent group’ sense never have a ‘commoner’ or ‘subordinate’ sense. A second distribution is also clear: although our evidence points to both ‘commoner, general populace’ and ‘attendants, subordinates’ meanings in Proto Nuclear Polynesian, Proto Ellicean and Proto East Polynesian, we observe that not a single living language retains both senses today. One or the other meaning has survived but never both. I take the ‘clan’ versus ‘commoner/subordinate’ distributions to be the result of geographical differences in meaning around Western Polynesia at Proto Polynesian times. The ‘commoner, populace’ versus ‘subordinate,

---

14 The East Uvean *kainanga form has a meaning suspiciously identical to that of Tonga and may be a loan. East Uvea was conquered by Tongans who came to reside there some centuries ago and the language has a heavy overlay of Tongan loans (Biggs 1980).
attendants’ distributions presently seem more serendipitous and I leave the question of interpreting those distributions to another day. Our present task is to point out that the ‘matrilineal clan’ sense had come to mean ‘commoners, populace’ to most Western Polynesian populations by Proto Polynesian society times.

As the extended quote from Marck (1996) in §3 indicates, I have felt since first examining the problem that the ‘commoner, subordinate’ senses emerged out of the ‘descent group’ sense. Three areas of progress in our general research now allow a more developed interpretation:

1. The emergence and current primacy of the Proto Oceanic society matrilineality model.
2. Abandonment of the notion that Proto Polynesian speakers necessarily had a single system of descent.
3. The observation that *kainaŋa senses of ‘commoner, subordinate’ are in complementary distribution when compared to the languages that have the ‘lineal descent group’ sense.

I now posit that the cognatic elements in the population existed through the middle and late Pre-Proto Polynesian period, with the result that the meaning of *kainaŋa was first transformed into ‘cognatic descent group’, then to ‘commoners of the cognatic descent groups’ and finally, by Proto Polynesian times, shifting to just the ‘commoner’ sense along the lines I suggested in 1996 (see §3 above). The evidence of the living languages is unanimous in indicating that a ‘cognatic descent group’ sense was no longer current in Proto Polynesian times and that only the ‘commoner’ sense existed.

The subpopulations retaining the ‘matrilineal clan’ sense would at least have included the ancestral interstage society which spawned the earliest Pukapukans and, perhaps separately, the ancestor of Anutan and Tikopian for which we must postulate a ‘unilineal descent group’ sense and, less certainly, a ‘matrilineal descent group’ sense. The ravages of time leave us with data that offer no compelling support for one interpretation over the other.

The situation is confounded by the possibility that there may also have been variability in local populations such as residually matrilineal seafaring folk in Tonga and Samoa and by our inability linguistically to specify a source for the origin of Pukapukan speech. Pukapukan is presently classified as a first order offshoot of Nuclear Polynesian and cannot be sourced to a particular place in western Polynesia but must have emerged out of Samoa, East Uvea or East Futuna rather than out of Tonga or the northern atolls (Tokelau or Tuvalu).

The source of lineality for Anutan and Tikopian is similarly muddied. Although I offered a family tree of Polynesian that ignored the Futunic Outlier group (Marck 1999, 2000), that tree was meant to show what kind of groups could be supported solely on the basis of exclusively shared sporadic sound changes. But such changes only occurred once every century or two in the history of the average living Polynesian language (cf. Marck 2000:59). Things were happening faster than that when Proto Nuclear Polynesian diverged into its primary subgroups (Futunic, East Uvean, Ellicean and Pukapukan). It was in some narrow band of time that the languages retaining the ‘clan’ sense diverged from the others. But we can be sure the ‘clan’ sense was current in the immediate ancestor of

15 ‘Pre-Proto Polynesian’ — that period between the divergence of Polynesian from all other Oceanic (circa 900 BC [Burley and Dickinson 2001; Burley and Connaughton 2007]) and break-up of unified Polynesian speech, a less certain time in the period AD 1–500.
Pukapukan, on the one hand, and Anutan and Tikopian on the other. In any event, the data force us to attribute a ‘matrilineal clan’ sense to at least some elements of both the Proto Polynesian and the Proto Nuclear Polynesian speaking populations.

5 Conclusion

By positing matrilineality (Hage 1998a, Marck 2008) and matrilocal systems (Hage and Marck 2003; Kayser et al. 2006) for the earliest Polynesians (circa 900 BC) and diversity in descent systems by Proto Polynesian times\(^\text{17}\) (circa AD 1–500), a coherent history for *kaināŋa can be proposed. Societies need not have homogeneous descent systems. We do not need, therefore, a unitary model of Proto Polynesian society’s descent mechanisms and the lexical data almost require that we do not. I posit a situation where the more sedentary elements of Proto Polynesian society may have been operating under cognatic principles while at least some groups, perhaps those most active in long-distance seafaring, were still matrilineal and perhaps still matrilocal. A reconstruction of ‘matrilineal descent group’ for Proto Polynesian *kaināŋa is indicated by the agreement between Pukapukan and Pohnpeic-Chuukic from Micronesia, but it need not have been the primary sense for most people, for whom a status term (‘[one’s] descent group’) seems to have become a rank term: ‘the general populace: commoners as opposed to chiefly families (*qariki)’.

Divale’s (1984) study suggests that societies which shift to matrilocal systems take an average of 1800 years to complete their return to patrifocal systems that have divested themselves of all formally matrilineal structures. This is roughly affirmed in the Polynesian instance where their ancestors in Melanesia were or became matrilocal in the period 1400–1000 BC and had, in the main, become cognatic and perhaps patrifocal in the period AD 1–500 in Polynesia. At that latter point in time, matriline persisted in some way — probably amongst groups who continued to experience long absences of men due to seafaring.

References


Biggs, Bruce, and Ross Clark, 1994, *Pollex: comparative Polynesian lexicon (computer data base)*.

---

\(^{17}\) In Marck (1996) I was writing solely as a linguist and attempted to take linguistic methods to their logical limit, specifically avoiding the fine details of social organisation questions which were, in any event, to remain nebulous in the Pacific Island social anthropology literature until two years later with the publication of Hage’s (1998a) matrilineality hypothesis.


Assaying the scope and number of fish names in Pendau based on Pawley’s Oceanic benchmarks

†PHIL QUICK

1 Introduction

The identification of fish names and other flora and fauna in a language is not a simple task. The challenges that a field linguist has in identifying and matching the scientific flora and fauna taxa with the folk taxa are complex. Any tool or method that can improve this task is welcome. Pawley (2004) has introduced a significant method for assaying the total number of fish names and the scope for which these can be systematically identified by fish family in Oceanic languages. I will call the statistics that Pawley introduces ‘benchmarks’ in this paper. A crucial part of these benchmarks is to distinguish between uninomial and binomial names (Pawley 2004; compare Berlin 1994, Bulmer 1992, and Osmond 2004). I will extend Pawley’s approach by applying boxplot statistics to his data and compare the Pendau (Sulawesi) fish taxa data to determine how complete the current inventory of 300 Pendau fish names is and determine where gaps may be. These gaps can be exploited to help complete the list of total names by going back to the field and focussing on those fish families where some names may have been overlooked. As discussed in §5, I have applied this to sharks and rays, and successfully increased the number of names.

2 Introduction to folk taxonomy conventions

Berlin (1992), Bulmer (1992) and Pawley (2004) show that it is likely that many, if not all languages utilise a similar naming convention for flora and fauna that parallels to some degree the biological scientific method. The folk taxonomy convention used here follows a similar pattern as used in the Linnaean system for ‘genus’ and ‘species’. For folk taxonomy the label ‘folk generic’ is used as a parallel to ‘genus’, and ‘folk specific’ is used as a parallel for ‘species’, as in (1). Above this level there are life-form names such as

---

1 An early portion of this paper was presented as a part of Quick (2006). Pendau is in the Tomini-Tolitoli language group in Central Sulawesi, Indonesia. Pendau speakers live mostly along the coastal area of western Central Sulawesi’s isthmus (Quick 2003).

2 For details of the methodology of this study, and boxplot statistics in particular, see §4.

621
‘fish’, ‘tree’, or ‘bird’ (which may also be used in a folk generic sense), as in (2). As a result of these conventions, it is standard to find uninomial names and binomial names. The former may be single or compound words that can identify a particular taxon, and the latter are always compound terms in which one component modifies or selectively identifies a particular taxon found in a folk generic group (two or more taxa).

1) Pendau folk generic
   mangiban ‘shark(s)’
2) Pendau folk specific
   mangiban bi’ung ‘hammerhead shark’

2) Sample Pendau life-forms
   bau ‘fish’
   puu ‘tree(s)’
   manu ‘birds (and chicken)’

3 Pawley’s benchmarks based on continuity, stability and innovations

Pawley’s research on patterns of stability and change in fish names (2004) sets some new benchmarks for lexicographers and provides some new tools for research in flora and fauna for coastal Austronesian languages. Pawley (2004) suggests that a minimum goal for the total number of fish entries should be no less than 300 taxa and a target of 400 taxa should not be unreasonable for languages of Oceania (with a likely extension of up to 500 names). Since the number of fish species in Indonesia and adjacent areas far exceeds that found in Oceania, it is not unreasonable that Pawley’s expectations in regard to Oceanic languages may apply to Austronesian languages in Indonesia and neighbouring countries, and thus these figures should be our minimal goals as well. For later comparison with the Pendau data, Table 1 presents Pawley’s data on the range of total recorded fish names found for nine Oceanic languages.

<table>
<thead>
<tr>
<th>Language:</th>
<th>Wayan</th>
<th>Marovo</th>
<th>Satawal</th>
<th>Gela</th>
<th>Palauan</th>
<th>Titan</th>
<th>‘Uvea</th>
<th>Kapina</th>
<th>Marquesan</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of names:</td>
<td>484</td>
<td>400+</td>
<td>400</td>
<td>368</td>
<td>336</td>
<td>287</td>
<td>284</td>
<td>262</td>
<td>260</td>
</tr>
</tbody>
</table>

Pawley’s research (2004) generally supports previous research by Bulmer (1992) and Berlin (1992) that uninomial names are more stable and more likely to be names that can be reconstructed, whereas binomial names are rarely reconstructable and can usually be demonstrated to be innovations. One important aspect of the research in lexicography as well as of comparative historical research is understanding whether the list of fish taxa obtained is complete or representative of the total names possible. Pawley demonstrates that a comparison of the data across related languages of the total number of fish taxa, as well as a determination of the ratio found between uninomial names and binomial names, can alert field researchers to whether their inventories have been complete or representative.

Pawley demonstrates that in the Oceanic languages about 30% of the fish names are typically binomial names, and he suggests that the ratio is ‘about 10 to 4 or 10 to 5 in favour of uninomial’ (2004:11). For the various Oceanic languages that he analysed, he found that the percentage of binomials ranged between 27% and 36%.

An exception to this rule in Pendau is that some reduplicated words are innovations, and technically these are uninomials. Other reduplicated words may not be innovations.
### 4 Assaying the scope and number of Pendau fish names

In this section I identify the total number of fish names in Pendau according to several categories. In the counts that I list for Pendau fish taxa in (3) I follow Pawley’s procedure (2004). Each name given for a taxon is counted, including synonyms and names for the different growth stages. Note also that a uninomial name can be more than one morpheme when the meaning of the morphemes is idiomatic (Pawley 2004:5). The lists given in (3) necessarily have some redundancy. In particular, six of the fifteen ‘freshwater fish’ names are also included in the total number of saltwater fish since they are ‘brackish water fish’, therefore subtracting nine freshwater fish names from the total number of fish taxa results in 291 saltwater fish taxa. Similarly, the tally for brackish water fish is included in both the tallies for saltwater fish and for freshwater fish. The adjusted total for saltwater fish taxa subtracts 12 ambiguous names, the reason for which will be explained below.

(3) Total Pendau Fish Taxa: 300
   Number of saltwater fish taxa: 291 (Adjusted total = 279)
   Number of freshwater fish taxa: 15
   Number of brackish water fish taxa: 6
   Number of taxa for whales, dolphins and dugong\(^4\): 4
   Number of total unambiguous uninomial names: 205 (= 73%)
   Number of total unambiguous binomial names: 74 (= 27%)
   Number of synonyms: 18 (37 total fish names)
   Ratio of adjusted uninomial to binomial names\(^5\): 3:1
   Ambiguous entries: uninomial = 3 and binomial = 9

In counting the total taxa there are four categories that I found I needed to follow in order to systematically count the taxa as they have been organised in my lexical database. These categories and their tallies are summarised in Table 2.

<table>
<thead>
<tr>
<th>Type of dictionary entry</th>
<th>Simple</th>
<th>Type A</th>
<th>Type B</th>
<th>Type C</th>
<th>Total</th>
<th>Adjusted total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninomials</td>
<td>187</td>
<td>29</td>
<td>—</td>
<td>—</td>
<td>216</td>
<td>(72%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>205 (73%)</td>
</tr>
<tr>
<td>Binomials</td>
<td>28</td>
<td>—</td>
<td>52</td>
<td>4</td>
<td>84</td>
<td>(28%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>74 (27%)</td>
</tr>
<tr>
<td>Total</td>
<td>215</td>
<td>29</td>
<td>52</td>
<td>4</td>
<td>300</td>
<td>(100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>279 (100%)</td>
</tr>
</tbody>
</table>

The first category, ‘Simple’ refers to entries that have no subentries or multiple groupings that complicate their classification. An example is given in (4).

(4) **Simple dictionary entry**


---

\(^4\) The classifier used for fish, *bua*, is also used for whales and dolphins. Thus, whales and dolphins are considered to be fish in the Pendau folk taxonomy.

\(^5\) Compare with Table 2 in Pawley (2004).
Type A reflects a category in which there is a uninomial Pendau fish name which has at least two subentries in which there are two or more transparent groupings of fishes by their scientific names. The example in (5) for *badonga*, shows the first subentry grouping two species of gobies, and the second subentry grouping a number of blennies. When all the subentries are listed as having the same uninomial name, this is counted as only one name in the tally.

(5) **Type A dictionary entry**

*badonga*

1) ~ sailfin ribbon-goby, robust ribbon-goby *Oxymetopon typus, Oxymetopon compressus* (Kuiter 1992:214)

2) ~ red-spotted blenny, banded blenny, breast-spot blenny, black blenny, pink-spotted blenny, horned blenny *Salarias, Salarias fasciatus, Salarias guttatus, Salarias fuscus, Exallias brevis, Parablennius intermedius* (Kuiter 1992:239–240)

Type B reflects a category in which there is a binomial Pendau fish name which further defines a uninomial Pendau fish name. For example, in (6) the *atamba’* is a main entry with two subentries. The first subentry lists the uninomial *atamba’* as various types of emperor fishes, and so this entry is counted as ‘Type A’. The second subentry lists the binomial *atamba’ banang* ‘striped emperor *Lethrinus ornatus*’ and is counted as ‘Type B’.

This example illustrates a case of including two names in the total inventory tally, one for Type A and Type B respectively.

(6) **Type B dictionary entry**

*atamba’*

1) ~ black-blotch emperor, orange-finned emperor, orange-stripe emperor, Lancer emperor, sky emperor, red spot emperor, tail-saddled emperor *Lethrinus harak, Lethrinus erythracanthus, Lethrinus obsoletus, Lethrinus genivittatus, Lethrinus atkinsoni, Lethrinus lentjan,* *Lethrinus erythropterus* (Kuiter 1992:82)

2) *atamba’ banang* striped emperor *Lethrinus ornatus* (Kuiter 1992:82)

Type C reflects a miscellaneous category for which none of the other counting methods works (and in my data only accounts for four fish taxa).

There were also a small number of ambiguous examples that were not counted (see Quick 2006, and forthcoming for details). By ambiguous I mean that there are some names which may be either a variation on a listed fish name or a nonce binomial consisting of a uninomial and a descriptive term that is not conventionalised. The twelve total ambiguous cases can be subtracted from the total in Table 2 above, along with a subtraction of the freshwater fish names (at least those which are not ‘brackish water fish’) to adjust the total to a conservative tally that subtracts 21 from 300, thus resulting in an adjusted tally of 279 saltwater fish names.

---

6 Entries such as *songko bolong* ‘various hogfish, and the juvenile half-and-half wrasse’ and *songkorong* are likely to be variations of the same taxon (the latter appears to be a contraction of the compound name), and thus are only counted as ‘one’ fish name. Another example is *buade* ‘various goatfish’ and *buluade* identified as a particular type of goat fish, that is the ‘blackbanded goatfish’. Other variations are clearly minor variants, such as *belu-belu* ‘various wrasses’ and *mbelu-mlbelung*.

7 The freshwater names are subtracted in order to give a better comparison with the Oceanic data that Pawley (2004) gives.
The 27% given as the adjusted total of Pendau binominal names in Table 2 is comparable to the figure Pawley gives for the Oceanic languages (between 27–36%). Although my research shows such interesting similarities to the statistics presented in Pawley (2004), it will be necessary to conduct similar analyses in a number of western Austronesian languages first before we attempt to make definitive claims or comparisons in case my results are coincidentally similar.

**Table 3:** Comparison of the number of Pendau fish names, family by family, with Oceanic fish names

(Oceanic data adapted from Pawley 2004:12–17, Tables 3–14; starred items indicate that Pawley has either recorded zero or has questionable data for some languages)

<table>
<thead>
<tr>
<th>Family</th>
<th>Pendau</th>
<th>Average Oceanic</th>
<th>Lowest Oceanic</th>
<th>Highest Oceanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serranidae – grouper, rock cod, coral cod</td>
<td>17</td>
<td>24</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>Carangidae – trevally and jacks</td>
<td>13</td>
<td>21</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>Labridae – wrasses</td>
<td>10</td>
<td>19.6</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>Scaridae – parrotfish</td>
<td>5</td>
<td>17.4</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>Mullidae – goatfish</td>
<td>10</td>
<td>11</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Mugilidae – mullet</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Scombroidae – mackerel and tuna</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>Xiphiidae – marlin, swordfish and sailfish</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Acanthuridae – surgeonfish and unicornfish</td>
<td>18</td>
<td>20</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>Siganidae – rabbitfish</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Balistidae – triggerfish</td>
<td>16</td>
<td>15.6</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Holocentridae – squirrelfish and soldierfish</td>
<td>10</td>
<td>10.4</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Lutjanidae – snapper, sea-perch, bass</td>
<td>17</td>
<td>14</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>Lethrinidae – emperors</td>
<td>10</td>
<td>11</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Haemulidae – sweetlips and grunters (Plectorhynchidae and Haemulinae)</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Sharks</td>
<td>12</td>
<td>11</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Rays</td>
<td>9</td>
<td>6.4</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Eels</td>
<td>8</td>
<td>11</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Sphyraenidae – barracuda</td>
<td>3</td>
<td>4.6</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Clupeidae – sardines and herrings</td>
<td>—</td>
<td>4.4</td>
<td>1*</td>
<td>7</td>
</tr>
<tr>
<td>Chanidae – milkfish</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Atherinidae – silversides and hardyheads</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Scorpaenidae – scorpionfish</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>Synanceiidae – stonefish</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Blennidae – blennies</td>
<td>2</td>
<td>2.5</td>
<td>1*</td>
<td>22</td>
</tr>
<tr>
<td>Gobidae – gobies</td>
<td>4</td>
<td>7</td>
<td>1*</td>
<td>23</td>
</tr>
<tr>
<td>Periophthalmidae – mudskippers</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Table 3 provides a tabular comparison of Pendau fish names family by family (or by similar families), following Pawley’s Oceanic data. This table provides a useful preliminary picture of the details we will examine more closely with boxplot statistics. Putting Pawley’s Oceanic data into boxplots provides a statistical method by which we can compare the Pendau data with the Oceanic data. The boxplot statistical method indicates the distribution of 50% of the data and the median point. Using boxplots provides a more robust and accurate method than the traditional statistical method of simple averages, means and deviations. By superimposing the Pendau data onto the Oceanic data in the boxplots we can highlight where gaps may occur in the current inventory of Pendau fish names. I assume here for the sake of comparison that the Pendau data for a particular fish family is likely to be low if it falls below the boxplot. Second, the median value can also be taken as a minimum target for determining how well the Pendau data correspond to the Oceanic data.

There is some evidence that the Pendau data correspond at least in some limited respects with the Oceanic data that Pawley (2004:13–16) has presented. This can be seen by taking Pawley’s data, in which he compares the number of fish names in eight Oceanic languages and putting them into boxplots. This is done for his most complete data in Figures 1 and 2 below (from Pawley’s Tables 3–11). The ‘x’ inside each boxplot marks the median value, and the ‘whiskers’ (lines above and below each boxplot) show the closest adjacent value outside of the interquartile range. Pawley also gives an average which I will note beside each boxplot with an ‘O’ (for Oceanic average) and arrow. His average only includes data from the five languages with the highest and most complete numbers of fish names. In the boxplots below I include all eight languages, as I am comparing the total number of fish names by family with Pawley’s Oceanic data and it is not yet known how complete the tally of Pendau fish names is. However, the nature of boxplot statistics removes outliers and spurious data (these are not shown in the graphs I give here as they are not relevant to my discussion), so for the most part the data below allow us to make a good judgment as to whether the tally of Pendau fish names for a particular family (or family groups) is on target or not. I will mark the Pendau total with a ‘P’ and an arrow.

---

8 ‘When summarising data numerically we will usually report the mean as a measure of location and the standard deviation as a measure of the spread of the data. We can in addition report the maximum and minimum values which gives the range of the data. If we pursue such an approach graphically we produce a boxplot. Boxplots are based on plotting the percentiles of data. The pth percentile is the value x such that p percent of data is less than x. For instance the 50th percentile is the value x such that 50% of the data is less than x . . . . . . . The components of the boxplot are interpreted as follows. The box displays the interquartile range, i.e. the distance from the 25th to the 75th percentile. Fifty percent of the data falls in this range. The whiskers are drawn to the nearest observation not beyond 1.5 times the interquartile range from the quartiles. Any points outside this range are plotted individually. The boxplots lets us see the location and spread of the data as well as assessing the skewness of the distribution . . . ’ (Cunningham, Barry and Walsh 1999:14)
Assaying the scope and number of fish names in Pendau

Figure 1: Boxplot of Pawley’s Oceanic fish data (2004) by families in comparison with Pendau (for groupers, rock cod, coral cod, trevally, jacks, wrasses, parrotfish, goatfish, mullet, mackerels, tuna, marlin, swordfish, surgeonfish, and unicornfish)

In Figure 1 the Pendau data are below the boxplot ranges for the Scaridae, Mugilidae, and Scombridae families. These indicate that Pendau is likely to have more fish names than have thus far been recorded, although the greatest discrepancy appears to be for the Scaridae family. Note, though, that a comparison of the raw data in Table 3 for these three families shows that these low tallies of fish names correspond to low tallies in the languages that Pawley examined. All of the other families in Figure 1 show that the number of Pendau fish names in each family compares favourably with the statistics for the Oceanic languages, although this analysis does not rule out the possibility that more Pendau names may be found.

In Figure 2 the Pendau names are all within the box plot ranges or higher, and generally have a medium to strong correlation with the Oceanic data. For the most part, the Pendau data corresponds to the Oceanic data and has a similar spread of representations in the fish families to those which Pawley recorded. This brings us to some final questions and observations.

One of the few continuities that exists between Oceanic and western Austronesian languages is in the taxonomy of marine flora and fauna. Since it appears likely that Pawley has discovered a norm which is applicable to Oceanic languages, the obvious question to pursue is whether this norm is also applicable to western Austronesian languages, as the Pendau data initially tentatively suggest. Part of the research needed to determine this includes assessing whether or not the higher number of fish species found in the Indonesian archipelago region will require an adjustment to Pawley’s Oceanic benchmarks, or if, as Pawley (2004) suggests, what may turn out to be more relevant than actual numbers of different species is the folk demarcation of similar groupings of fish according to their degree of importance, based on considerations such as economic benefit, usefulness, danger, unusual characteristics and so on.
Further aspects that may require fine tuning include specific contexts for languages with other complicating factors such as language contact. Pendau will serve to illustrate whether or not the total number of fish names can be compared directly with the norm that Pawley has identified. Contact with other languages over time may have interfered to some extent with the retention, identification and renaming of fish names. These languages include the local lingua franca Kaili, as well as Bajau communities in the region, and the influence of Indonesian more recently. Other long term events may have contributed to the loss of some fish names, in particular the raiding and enslavement of the Pendau that occurred for a period of time and contributed to the Pendau having less exposure to the sea and more to the jungle environment. The Pendau are traditionally swidden agriculturists as well as being strongly involved in hunting and fishing. Future research will need to determine whether or not a language group under investigation needs to be a ‘complete’ fishing society in order to be usefully compared with Pawley’s benchmarks, or if they can be profitably applied to a community with a mixed economy which includes a strong fishing component, as the Pendau data indicate to be the case.

5 Utilising Pawley’s benchmarks to strategically search for fish names

In this section I discuss how a comparison of the Pendau and Oceanic data led to adding nine new fish names to the Pendau inventory for sharks and rays. The data in Figure 3 repeats a portion of Figure 2, specifically that relating to sharks and rays. As before, the O indicates the Oceanic average. In addition, the P followed by 2005 indicates the tally of names as of January 2006 (Quick 2006), that is, collected prior to 2006. The P followed by 2006 indicates field research conducted in 2006 that focussed on identifying new names for sharks and rays. In that follow-up research I discovered five new shark names and four new ray names as listed in (7) and (8) respectively.

Figure 2: Boxplot of Pawley’s Oceanic fish data (2004) by families in comparison with Pendau (for rabbitfish, squirrelfish, soldierfish, triggerfish, leatherjackets, snapper, sea-perch, bass, emperors, sharks, rays, eels and barracuda)
Figure 3: Boxplot of fish name tallies before and after gaps were identified

(7) New shark names added
- ulunan alibambang ‘angel shark, Squatina sp.’
- mangiban balida ‘thresher shark, Alopias pelagicus’
- mangiban lalua ‘whitetip reef shark, Triaenodon obesus’
- mangiban manasap ‘nurse shark, Ginglymostoma cirratum’
- mangiban songka ‘bronze whaler shark, white shark, Carcharhinus brachyus, Carcharhinus carcharias’

(8) New ray names added
- ulunan ‘giant shovelnose ray, Rhynobatos typus’
- pagi lana ‘southern stingray, Dasyatis americana’
- bole’ bambanga ‘manta ray, Manta birostris’
- bole’ bulalu ‘reticulate whipray, Himantura uarnak’

The procedure I followed was to take additional references to the field which included more and better photographs and/or pictures of sharks and rays (e.g. Sharks and rays of the world, Perrine (1999) and Sport fish of the Pacific, Dunaway (2004)). Previous references I used had either no pictures of rays or sharks or were very limited as to the range of the different species. Pendau assistants looked through the pictures and identified those fish they were familiar with. This process led to adding the nine new fish names to the Pendau inventory.

6 Conclusion
Although this research is preliminary, the fact that the languages examined are all Austronesian and that there is continuity of marine fauna classification between these
linguistic groups provides grounds for suspecting that the statistical findings are not merely coincidental. The importance of the continuity of marine fauna found in the Indo-Pacific regions, along with the important demarcation between binomial and uninomial names, is discussed by Pawley (2004:19):

The dramatic difference between uninomial and binomial retention rates, presumably, has something to do with the geographic distribution of species as opposed to genera. Folk generics, represented by uninomials, are typically applied to whole families. The same genera and families of fish tend to occur throughout the tropical Indo-Pacific. By contrast, folk specifics, often represented by binomials, typically apply to a single biological species or a group of closely related species and the geographic distribution of species tend to be more localised. Thus modifying terms for folk specifics do not ‘travel’ as well as generic terms. But perhaps the main reason is that while generic names are usually arbitrary, modifiers almost always describe a feature of the morphology, the behaviour or the ecological niche of particular species. Accordingly, a number of competing modifiers may suggest themselves as equally convenient ways of distinguishing a particular member of a folk genus, so that even when a species is ubiquitous the original modifier in a binomial will often be subject to competition.

Bulmer (1974:11–12) also makes some interesting comments about the continuity and discontinuity of the ecological environment in relationship to the total inventory of taxa names:

First one must appreciate the geographical limits of the ecological space within which individual human communities, and indeed individual human naturalists, operate. Any one preliterate human community occupies only one tiny fragment of the earth’s surface — and also, even with the extension backwards of its oral tradition, one brief moment of time. Thus the total number of animal and plant species with which it can have significant relations is only a minute proportion of the total number of animal and plant forms accessible or potentially accessible to modern science. While in global space and time all animate life is continuous, restrictions of space and time at once impose discontinuities, and in the micro-space occupied by a single human community, or a set of related communities, objective discontinuities abound. This is not solely because of the reduction in total number of species present, but because competition for ecological niches restricts the number, in any limited area, of closely similar forms.

Pawley’s Oceanic benchmarks provide lexicographers with a new tool for assaying a folk taxonomy of fish names in Austronesian languages. This provides us with a means to determine whether a particular inventory tally is representative and/or nearing completion. Although lexicography is notoriously an open-ended field of research, the idea that a particular domain can be established as relatively complete is a significant landmark. Other clear benefits include establishing a quality inventory that will be of great benefit to researchers, especially in the fields of historical comparative linguistics and ethnobiology. Pawley’s Oceanic benchmarks pave the way to bringing further payoffs as his method is adapted and perhaps refined in relation to other Austronesian languages.

References


—— forthcoming, Fish names inventory: insights into the Pendau language and culture. *Journal Masyarakat Indonesia (Sulawesi edition).*
1 Introduction

This paper looks at some of the common metaphors used in Tetun Terik, an Austronesian language of eastern Indonesia, to talk about dispute resolution. As indicated by the title, it draws heavily on the insights of Lakoff and Johnson’s book *Metaphors we live by* (1980). They note that many of the ways in which we speak about the world are metaphorical. This is not just a matter of language; rather, the language reflects the fact that we view certain mental domains in terms of others. Thus these metaphors provide a key to discerning how the speakers perceive reality. For instance, in English we frequently speak about arguments in terms of buildings (e.g. one can ‘construct a good case’, or argue that ‘your claims have no foundation’), or alternatively in terms of war (e.g. one can ‘shoot down an argument’ or ‘defeat an opponent’). The first presents a careful step-by-step approach to argumentation, the latter an interactive, adversative one. In this way, each metaphor highlights a particular aspect of a concept. At the same time, the metaphor obscures alternative views. So, for example, the war metaphor obscures the possibility of working cooperatively to resolve conflict.

We will see that the metaphors used for dispute resolution in Tetun Terik are very different to those used in modern court-based systems, and that these metaphors reflect very different goals and emphases. Indeed, one of the challenges faced by the legal systems in both Indonesia and East Timor is how to best incorporate two very different systems for handling disputes, namely the traditional ones, and court-based ones.  

Data for this paper was collected during fieldwork in 1993 and 1995 within the ancient kingdom of Wehali, centred around the town of Betun, in the south of the district of Belu in West Timor, eastern Indonesia.² This community speaks the Fehan dialect of Tetun.

1 The relationship between the two systems of handling disputes is a recognised issue in legal circles in both Indonesia and East Timor. It was one of the themes of *The International Conference on Traditional Conflict Resolution and Justice* (Dili, East Timor, 27th June 2003), at which an earlier version of this paper was presented.

2 Professor Andy Pawley was one of my PhD supervisors for this research, in the Research School of Pacific and Asian Studies, The Australian National University. At the time I was, like most post-graduate students in his department, preoccupied mainly with grammar. Over the succeeding years, as I have sought to acquire ‘nativelike fluency’ in Tetun, I have increasingly responded to Andy’s urging to take an interest in words (e.g. Pawley and Syder 1983; Pawley 1992).
Terik (van Klinken 1999), a language which is also spoken by some communities over the border in East Timor. Formal Tetun Terik oratory makes heavy use of poetic metaphors (Therik 2004:39). This article, however, draws primarily on everyday talk about dispute resolution, with only occasional reference to the ritual register.

2 Mediators are bridges

In Wehali, the term *ai kalete* (wood/plant bridge), or simply *kalete*, literally refers to a bridge. It is also used metaphorically to refer to mediators in disputes, as well as the go-betweens who mediate between the man’s family and the woman’s family in traditional courtship and marriage negotiations. This role is described briefly in example (1), and in more detail by Therik (2004:102ff.).

(1) If a man has been courting a girl, and he likes her, he talks with his parents to ask for her hand in marriage.

*Haruka ai.kalete. Foin mai horan hak-ee*

send bridge then come sense say-HES

‘(They) send a mediator. (She) then comes to feel out (the situation) saying,’

*“O-oan sia mai labu ne’e, emi haho’uk ka la ho’i?”*

RDP-child PL come court this 2p agree or not don’t.want

“The children coming here to court, do you agree, or disapprove?”

*Haho’uk ti’an, at ami la’o hodi mama oan ida mai.”*

agree already IRR 1pe walk bring betel small one come

(If) you agree, we’ll come, bringing a little betel.”

What is the basis for comparing mediators to bridges? Here the hypotheses are necessarily somewhat speculative. As Lakoff and Johnson point out, only part of a metaphor is typically ‘used’, with other possible bases of comparison being ignored. The Tetun metaphor seems to be based on two similarities, namely that there is a separation between the two sides, and that somebody or something links the two sides despite their separation.

In Wehali, a mediator is a valuable asset in bringing two sides of a dispute together. Without some initial approach between the parties (whether through a mediator or in person), no resolution can be reached.

3 A legal case is serious talk

The Tetun Terik word *lia* has a wide range of meanings. It refers to any unit of language, including languages (e.g. *lia Tetun* ‘the Tetun language’), words, sentences, and news. It indicates serious talk, rather than just chatting to pass the time (Therik 2004:39). *Lia* also refers to traditional communal events such as weddings, which involve serious negotiation. Within the legal domain, it refers to court cases and legal disputes, regardless of whether these are handled traditionally by the elders, or in less traditional ways by the police or court. Both are illustrated in example (2).

(2) Recently there was a problem in our village about one person’s pig entering someone else’s garden and destroying the corn crop. The garden owner retaliated by killing the pig.
So the owner of the garden and the owner of the pig had a dispute.

The problem escalated to the garden owner hitting the wife of the pig’s owner. The garden owner did not accept the verdict of the traditional elders that he should pay a fine for dishonouring the woman.

Then the garden owner and others (i.e. his wife and friends) took this case to the police.

The police did not resolve it either, so the case was brought back to the traditional elders, as described in example (5).

Here the basis for comparison appears to be that resolving disputes crucially involves serious talking. Perhaps there is the further implication that such talking will typically persist over a long time and involve many people, just as in weddings and their associated inter-family negotiations.

4 Judging is weighing

There are at least three Tetun Terik expressions that appear to be based on the underlying metaphor of judging as weighing.\(^3\)

An everyday example is the use of tetu lia (weigh word) to express ‘judge a court case’. Tetu literally means ‘weigh’; it is for instance the word used for weighing babies, and for weighing buffalo for sale. It also extends to carefully considering a situation, and to choosing one’s words carefully, anticipating their impact.

(3) When we want to reconcile with an enemy:

\[
\text{Ita tetu lia.fuan ne’e, bat kona ba ema nia-kan neo-n.}
\]

1pi weigh word this so that touch to person 3s-poss inner.nature-gen

‘We weigh up (our) words so that they’ll touch the person’s conscience.’

A second term based on the metaphor of judging as weighing is dasin. On a literal level, a dasin is a set of scales. Such scales are traditionally of the hanging type, where the item to be weighed is put on one arm, and weights are placed on the other arm until the two are level. Within the ritual register of Tetun Terik, dasin also refers to nobles and their appointees. One man explained this by saying that such people are able to ‘weigh’ court cases.

---

\(^3\) My data does not include examples of weighing a crime to determine its seriousness, or of weighing evidence to determine its validity. In church use, however, wrongs are spoken of as ‘heavy’ or ‘light’ (Tom Therik pers. comm.; Genesis 9:25 in Mathijisen (1967)). Such expressions are also common in modern Tetun Dili, the Tetun-based lingua franca of East Timor, which is strongly influenced by Portuguese. There, for instance, krimi todan (crime heavy) is one way of expressing the notion ‘serious crime’.
Thirdly, there are *ktetuk* ‘level, flat’ (derived from *tetu* ‘weigh’ by the adjectival circumfix *k*- ~ *-k*), and *nesan* ‘level, same’. These words are frequently paired together, particularly within ritual Tetun, which uses much poetic parallelism. Metaphorically, these terms refer to reaching agreement, as illustrated by the following accusation against those who judged a case.

(4) *Emi-kan ha-kotu lia la ktetuk, la nesan.*

2p-POSS make-severed word not level not level/same

‘Your judgment wasn’t ‘level’ (i.e. one side disagreed with it, or it wasn’t fair).’

The metaphor of judging as weighing is of course familiar to English speakers, and to all who accept the symbol of scales as a symbol of justice. According to Wood (2005), ‘the scale as a metaphor for justice equates balance with fairness and imbalance with unfairness and mistreatment.’ In Tetun Terik, however, the goal of weighing appears to be agreement by both parties, rather than the more impersonal notion of both parties being treated fairly.

5 Giving a verdict is severing

Reaching the final verdict in a case is commonly spoken of in terms of severing. There are several expressions based on this metaphor. These include three morphologically related words: the intransitive root *kotu* ‘be severed’, transitive *hakotu* ‘sever, cut off’, and the derived abstract noun *kakotun*.

Intransitive *kotu* means ‘be severed’. For instance, the nobleman Suri Tuan in one story suffers the fate of having someone *taa kotu ulun* (chop severed head) ‘cut off his head’. Metaphorically, the verb means ‘be finished, be concluded’. So, when a story-teller was interrupted by a woman who thought she had finished, she could protest, *Seidaun kotu, ina!* (not.yet severed mother) ‘It’s not finished yet, ma’am!’. When buyer and seller agree on a price, one can say that *folin kotu* (price severed) ‘the price is agreed’. Within legal disputes, when one says *lia kotu* (word severed), it means that the case has been concluded.

The transitive verb *hakotu* (make.severed) means to chop right through something, such as a piece of wood, severing it. *Hakotu* metaphorically extends to reaching a final decision. In economic transactions, *hakotu folin* (make.severed price) refers to reaching an agreement on price. Within the context of legal disputes, *hakotu lia* refers to reaching the final decision that will conclude the case. That is, it equates to *halo lia ne’e kotu* (make word this severed) ‘conclude the case’. A similar expression is *ko’a lia*, with *ko’a* literally meaning ‘cut’ (for instance, with a knife).  

(5) Manek, accused of hitting his neighbour’s wife, did not accept the fine which the police wanted to impose on him, because that fine included payment of money.

*Dadi lia ne’e sia r-ean hikar mai iha adat,*  
so word this 3p 3p-pull back come LOC tradition  
‘So they dragged this case back to the traditional leaders,  
*folin r-oo malu r-a-kotu.*  
only.then 3p-accompany each.other 3p-make-severed  
and only then did they together conclude (the case).

---

In Tetun Dili, the Tetun-based lingua franca of East Timor, this expression has a much more general meaning, as simply ‘speak’. However, Tetun Dili retains the metaphor of giving a verdict as severing in *tesi lia* (chop/sever word) and *hakotu lia.*
Metaphors we judge by: mediation in Wehali

R-a-kotu tuir Manek nia-kan hakara,
3p-make-severed follow Manek 3s-POSS like
They concluded it in accordance with Manek’s wishes

nia simu n-ola sala ne’e, …
3s receive 3s-take wrong/fine this
He accepted the (imposition of the) fine, [namely, one pig,
four bottles of wine, and one hand-woven cloth].’

Kakotun, the abstract noun derived from kotu ‘severed’, can similarly be used metaphorically to refer to a final decision in a dispute, as in the following report.

(6) Lia ne’e nia-kan ka-kotu-n, feto fahi ida, foos karon ida,
affair this 3s-POSS NR-severed-NR female pig one husked.rice sack one
‘The final decision was that the woman (had to give) a pig and a sack of rice,
mane moos nu nia.
mane also as 3s
and the man had to do the same.’

What is the basis on which reaching a verdict is compared to severing? The evidence presented above, that kotu also means ‘be finished’, suggests that the comparison is based on finality. A successful verdict is one that is accepted by both parties, a point which is explicitly made by the speaker in example (8) in the next section. When such a verdict is successfully delivered, the dispute is finished, and one can say hakotu hola (make.severed take) ‘the decision succeeded’. If, as in example (7), it is unsuccessful, one can say hakotu la hola (make.severed not take) ‘the (attempt at) decision didn’t succeed’.

(7) After a young man refused to marry a young girl with whom he had had intercourse, the girl’s entire family took the case to the village head.

Baa, desa n-a-kotu la n-ola. Ami hoo malu hakees,
go village 3s-make-severed not 3s-take 1pe accompany each.other converse
‘(We) went, and the village head failed to conclude the case. We discussed it together:
kalol desa n-a-kotu sia la tuir, ami hodi tenik ba camat.
if village 3s-make-severed 3p not follow 1pe bring again to sub-district since the village head made a decision and they (the other side’s family) wouldn’t follow it, we would in turn take (this problem) to the subdistrict head.’

6 Modern court cases are competitions

As noted in the introduction, English frequently expresses disputes in terms of wars. I am not aware of any traditional expressions based on the war metaphor in the Fehan dialect of Tetun Terik. When speaking about modern court-based justice, however, this metaphor does creep in, at least to the extent that one party will manaan ‘win’, while another will lakon ‘lose’ or monu ‘fall’.

English, of course, also has a range of metaphors based on cutting. However these pick up on very different features of cutting. For instance, some are based on cutting being painful (e.g. ‘cutting remarks’, ‘he’s cut up about it’), some on cutting dividing one entity into two (e.g. ‘sever a relationship’), and some on cutting to remove unwanted parts (e.g. ‘cut a text’, ‘cut to the point’).
The case in example (7) was ultimately taken to the judge in the district court.

Iha nia, hakim n-a-kotu, n-alofeto maaan. ...

Loc 3s judge 3s-make-severed 3s-make female win ...

‘So, the judge concluded the case, making the woman win. [The judge concluded, making him pay 1,316,000 rupiah, a horse, a woven cloth, and a pig. The man’s family went along (with this decision). They accepted it.]

... dadi oras ne’e, n-osi mane monu, mane tate ta
... so time this 3s-from male fall male pay.fine already

ba feto ti’an,
to female already

[Having accepted, having responded.] so now, the man has fallen (i.e. lost), and has paid the fine to the woman.

hakim n-a-kotu, sia simu, sia tate, dadi moos.
judge 3s-make-severed 3p accept 3p pay.fine so finished

(In summary,) the judge has concluded the case, they accepted (the decision), they paid the fine. So it is finished.’

7 What is in focus?

The aspects of dispute resolution which are highlighted by the various metaphors discussed above may be summarised as follows: The ‘MEDIATORS are BRIDGES’ metaphor focuses on two parties being separated, but there being somebody who nevertheless provides a link between them. ‘A LEGAL CASE is SERIOUS TALK’ highlights that resolving disputes crucially involves talking. ‘JUDGING is WEIGHING’ underlines the importance of both parties accepting the decision. In the metaphor ‘GIVING A VERDICT is SEVERING’ we see that once the verdict is delivered and accepted, the issue is concluded.

Together, these metaphors emphasise that the goal of traditional dispute resolution is a restored relationship between two separated parties, with a verdict that is accepted by each. The focus is thus on moving on towards the future.

The above metaphors express an ideal. There are of course many disputes that don’t reach resolution in the traditional system. The parties may not even get to the stage of seeking a mediator, leaving the issue unresolved and the relationship broken. Then again, the parties may seek a solution without words, through means such as a revenge attack or sorcery. Finally, the parties may not accept the decisions of the elders, and may refuse to be reconciled. Nowadays, when there is a failure to reach a resolution, the dispute may be brought to the courts, for resolution in the ‘modern’ system.

6 Similar points were made by other speakers at The International Conference on Traditional Conflict Resolution and Justice (Dili, East Timor, 27th June 2003). For instance, President Xanana Gusmão expressed the goal of traditional conflict management in East Timor as being peace or reconciliation, while the Vice-Chancellor of the National University of East Timor, Dr Benjamin de Araújo e Côrte-Real, described it in terms of sustaining the social order.
8 What is obscured?

Lakoff and Johnson point out that metaphors not only illuminate, but they also hide alternative ways of looking at the world. For instance, the English 'COURT CASES are WARS' metaphor hides the possibility of seeking cooperative solutions. It is not that such solutions are impossible in English, but they are not supported by this metaphor.

It appears that a number of principles which are important in the court system of justice have little or no place in everyday talk about traditional dispute resolution in Wehali. Such concepts include the following.

- 'guilty' and 'innocent': The closest terms I have found are the very generic sala, which means 'wrong, incorrect, err, mistaken, sin, not according to the rules' (as well as being a 'fine' paid by the wrongdoer), and its negation la sala 'not wrong'. The adjective loos 'straight, true' can describe true statements, but does not appear to extend to innocent people.

- 'just' and 'fair': One option for approximating these concepts is loos 'straight, true', as opposed to kle’uk 'crooked' (Eugenio Sarmento and Tom Therik pers. comm.). Another appears to be ktetuk 'level, flat' and nesan ‘level, same’; however these paired words refer more to mutual agreement than to fairness as an abstract concept. Another option is kona ‘touch’. When someone’s speech is said to la kona ‘not touch’, it can mean that it is wrong, or that it is unacceptable in some other way, perhaps because it is not intelligible or not just. When people’s speech kona malu (touch each other), they agree.

- individual rights.

While it is possible that more exact terms for these concepts could be found by targeted elicitation, the fact that they appear to be absent from the author’s dictionary file of over 3000 head-words and her corpus of 64,000 words of spoken texts, suggests that they are, at best, not central to the way people speak.

9 Conclusion

We have seen that traditional Tetun Terik dispute resolution and the modern court-based justice system each have very different goals, insights, and blind spots, which are reflected in the different metaphors used to speak about them. They reflect two different cries frequently heard in East Timor in the aftermath of the violence of recent years: on the one hand the push for reconciliation, and on the other the plea for perpetrators of violence to be brought to justice.

One of the many challenges facing communities in both the Indonesian and East Timorese parts of the Tetun Terik speaking region is to seek ways in which they can benefit from the strengths of each system. Can one maintain the admirable goal of reconciliation when cases are brought before modern courts? Is there a risk that the overarching goal of restoring relationships in the traditional system will at times lead to an undervaluing of justice, resulting in festering sores which later lead to renewed disputes? When brought before the courts, do Timorese from traditional backgrounds understand that the goals and methods are so different to those found in traditional systems? Answering such questions could point the way to a more successful and locally relevant overall justice system for both Indonesia and East Timor.
References


1 Introduction

When compiling a bilingual dictionary for a relatively undescribed language one of the decisions that has to be made is whether or not to have entries for individual grammatical functor morphemes, many of which are likely to be affixes. The case for including entries for such morphemes rests primarily on the argument that the functions of a dictionary of this kind include fostering awareness of the grammatical structure of the language and awareness of how that structure works. Development of these awarenesses has the potential to be of value to both native speakers of the language and to those who, with various kinds of interest, come to it from the outside. The inclusion of dictionary entries for grammatical functor morphemes draws attention to their existence and points the reader towards information in the dictionary’s Introduction on how the grammar of the language works.

This paper considers the lexicographic presentation of verb phrase grammatical functors of Raga, an Eastern Oceanic language spoken by a total of around 6500 people, most of whom live in the northern third of Pentecost Island, Vanuatu.

1 It is a privilege and a pleasure to be contributing to a volume honouring and celebrating Andy Pawley’s four decades of academic achievement in research and teaching. He and I were student contemporaries in the Anthropology Department at the University of Auckland in the heady days of the early 1960s, and our paths have subsequently intersected in Port Moresby, Honolulu, Bali, Suva, London, Port Vila, Auckland, Wellington, Christchurch, Dunedin, Canberra and Sydney. In Honolulu he persuaded me into an enjoyable guest appearance for the many-nationed local cricket club, and our sabbatical presences in London coincided with our being able to savour an England v India Test at Lords.

Fieldwork and research underlying this paper were funded at various times and in varying proportions by The Australian National University, the Australian Research Grants Committee, the Myer Foundation, the University of Auckland and the University of Sydney.

The comments of the editorially commissioned reader of the paper are much appreciated — they pointed me towards several significant improvements.

My thanks go to Richard Leona, Raga co-compiler in the dictionary project, for invaluable insights into the verb phrase structure of his language. Responsibility for any misinterpretations of that system which may be present in the analysis rests of course with myself.

Richard, who has known Andy for around 25 years, joins me in wishing him well in his retirement.
1.1 Raga Orthography

The five Raga vowel phonemes — /a/ /e/ /i/ /o/ /u/ — and ten of the eighteen consonant phonemes — /d/ /h/ /k/ /l/ /m/ /n/ /r/ /s/ /t/ /w/ — have one letter representing one phoneme and their phonetic values are broadly conventional. Orthographic representations and brief phonetic descriptions for the remaining eight consonant phonemes are: /b/ a bilabial stop + voicing, /bw/ a bilabial stop + voicing + labio-velarised release, and /g/ a velar stop + voicing + homorganic prenasalisation; /v/ a labio-dental fricative + voicing, /vw/ a labio-dental fricative + voicing + labio-velarised release, and /g/ a velar fricative + voicing; /mw/ a bilabial nasal + labio-velarised release, and /n/ a velar nasal. For three stops, /b/ /bw/ and /d/, homorganically prenasalised allophones are obligatory after NV(V) where N = nasal consonant and V = vowel, with /b/ being + voicing when prenasalised. After N(V)hV(V) homorganically prenasalised allophones are optional for these stops.

2 The Raga verb phrase

There are two types of verb phrase, primary and serial. In its minimal form the primary phrase consists of an unmodified verb base used imperatively. More usually the active mood primary phrase consists of a verb base modified by various grammatical functors. Some of these functors are affixed to the verb base (see §2.1 and §4.4.1), while others, here termed peripheral, are not (see §2.2 and §4.4.2). A positive active mood primary verb phrase can be followed by one or more (seldom more than two) serial verb phrases (see §2.3 and §4.4.3). The passive mood primary verb phrase is considered in §2.4 and §4.4.4.

A member of a small subset of adverbs can be interpolated between the verb and the non-transformative suffixes in a primary phrase. This interpolation possibility is not given detailed and specific consideration here. Also not considered here are various quasi-verb constructions in the numeral system.

The semantically significant total or partial reduplication which can occur with many verb bases is a process rather than a functor, and it is therefore not considered here, but will be described in the dictionary’s Introduction, with reduplicated verb forms being included in the dictionary as derived sub-entries.

When allomorphic variation is present in cited data ~ indicates phonological conditioning (§4.1) or free variation (§4.3), and <==> indicates morphological conditioning (§4.2). Where a zero alternant is cited this indicates that in certain specified conditions a given grammatico-semantic function is signified by the absence of any overt marker.

Examples for the various functors listed in §2.1 through §2.4 are provided in §3.1.

2.1 Functors affixable to the verb in an active mood primary verb phrase

The prefixed functors and some of the suffixed functors transform the simple verb in various ways and in so doing create compound verbs.

The prefixed transformative functors are va- <==> ba- ‘causative’, and vi- <==> bi- ‘reciprocal’, which latter occurs only with wholly or partly reduplicated verb forms. Each of these prefixed functors can also occur in some contexts other than the verb phrase.

The suffixed transformative functors are: -ana ~ -na ‘passive’, 3 and -i ‘transitivising’.

---

The suffixed non-transformative functors are: -gi ‘unspecified object complement’, -ni-
<=-zero- ‘indirect object ligature’, and the paradigmatic set of object pronouns -au ~ -u
‘1s’, -go ‘2s’, -a ~ -e ‘3s’, and -ra ‘3p’. Each of these suffixed non-transformative functors
can also occur in contexts other than the verb phrase.

2.2 Functors peripheral to the verb in an active mood primary verb phrase

All but one of these functors are preposed to the verb base. Some of them are isolable
morphemes, and some only occur affixed to other functors. Two of them can also occur in
a serial verb phrase.

The preposed peripheral functors are: the paradigmatic set of non-isolable subject
the paradigmatic set of non-isolable number markers -ru ‘DU’ and -tol ~ -dol ‘TRL’;
the paradigmatic set of tense/aspect markers mwa <=> -m, -mu-,-m- with -m ~ zero ‘in progress
(present time)’, nu <=> -n with -n ~ zero ‘completed (past time)’, vi <=> -v with -v ~ -i
‘future’, si <=> -s with -s ~ -s- ‘conditional action or condition’, and men <=> -men
‘desiderative’; hav ~ -av the first morpheme of the discontinuous negation marker; siv ‘just’;
and mom ‘still/yet’. These last two functors can also occur in a serial verb phrase (see §2.3).

The only postposed peripheral functor is tehe ~ te the second morpheme of the
discontinuous negation marker.

2.3 Functors peripheral to the verb in a serial verb phrase

Preposed ba ‘serial verb phrase marker’, which is incompatible with a preceding
primary verb phrase containing hav ~ -av.... tehe ~ te, occurs only in a serial verb phrase.
Two peripheral functors listed in §2.2 above, siv ‘just’ and mom ‘still/yet’, can also occur,
with some constraints, after ba in a serial verb phrase.

2.4 Functors peripheral to the verb in a passive mood primary verb phrase

When a verb has been transformed to a passive form by -ana ~ -na it can be preceded
by whichever one of the tense/aspect markers mwa, nu, vi, si, men is semantically
appropriate in a given context. The possibilities of + hav....tehe and/or + siv and/or + mom
in phrases of this kind are under investigation.

3 Lexicographic presentation

For most users of the dictionary the starting point of their acquaintance with Raga verb
functors will be the individual entries. These are given below in alphabetical order, using
the standard entry typography.4

Each substantive entry includes the head-word or head-morpheme with allomorphs
where applicable, a brief ‘part-of-speech’ description, a referral to the relevant section or
sections of the grammatical information in the Introduction to the dictionary (not
reproduced here), a gloss and/or a statement of function, examples of usage (in which, in

---

3 Raga also has two gerundive transformative verb suffixes, -ana ~ -na and -va ~ -v which create compound
verbal nouns.

4 Abbreviations used in this paper are not exactly the same as those that are used in the dictionary.
the context of this paper only, hyphens indicate intra-word morpheme boundaries), a Bislama (BIS) equivalent where applicable, and cross-references where applicable.

Where a given functor has allomorphs the main criteria for determining which allomorph is to be listed as that functor’s main entry are, in descending order of application, under its isolable allomorph if there is one, under its most widely occurring allomorph, and under its longer affixed form if applicable.

Allomorphs other than those which are head-words or head-morphemes occur not only as alternant items following head-words or head-morphemes, but also in individual cross-referencing entries linking them to the appropriate substantive entry.

The entries do not include detail of the conditioning for allomorphs (§4), as these are matters described in the dictionary’s Introduction.

As mentioned in §2, allomorphic variation that is phonologically conditioned or free is indicated by ~ and variation that is morphologically conditioned is indicated by <=>. Where zero is cited as an alternant this signifies that in certain conditions the grammatico-semantic function in question is indicated by the absence of any overt marker.

In the examples in §3.1 there are instances — gita <=> ãita ‘see’, gani <=> ãani ‘eat’, and vano ~ van <=> bano ~ ban ‘go’ — of the morphologically conditioned variation of certain verb-initial consonants which is briefly described in §4.2 in relation to va- <=> ba- and vi- <=> bi-.

### 3.1 Dictionary entries for verb phrase grammatical functors

- **-a ~ -e**, object pronoun (3s). Him/her/it. Mwa ãobe-a He/she bends it (referring to bending a tendril-guiding reed used in cultivating damu, sivwiriği and bwevu). Na-v loli-a I will do/make it. Wai mwa asi-a Water cools him/her/it. Na-n vulu-a I stuck/joined it together. Na-m ãita-e I see him/her/it. Go-m roño-e You hear him/her/it. BIS hem.

- **-ana ~ -na**, passive transformative. Transforms a verb compatible with the active mood to one compatible with the passive mood. Nu gani-ana na damu The yam was eaten. Mwa ãani-ana na damu The yam is being eaten. Vi gita-na na vusi The hill will be seen.

- **-au ~ -u**, object pronoun (1s). Me. Ra-n roño-au They heard me. Go-m ãita-u You see me. BIS mi.

- **-av**, see hav.

- **ba**, serial verb phrase marker. And. Ra ban ba hae they are going and going up. Na-i van ba gita-e I will go and will see him/her/it.

- **ba-**, see va-.

- **bi-**, see vi-.

- **-dol**, see -tol.

- **-e**, see -a.

- **ga-**, subject pronoun (1p.excl.). We (more than three people excluding addressee/addressees, or ref. two or three people excluding addressee/addressees in a context in which number has already been specified). Ga-n van We (exclusive) went. BIS mifala.
ge- <=> zero-, subject pronoun (3s). He/she/it. Bere ge-m lai-a Lest he/she takes it. Nu gita-go He/she saw you. BIS hem.

gi-, subject pronoun (2p). You (more than three people, or ref. two or three people in a context in which this has already been specified). Gi-v gita-e You (plural) will see him/her/it. BIS yufala.

-gi, unspecified object complement. Indicates an unspecified object. Ra-v tutui-gi They will hit (something unspecified).

go-, subject pronoun (2s). You. Go-v gani-a You (singular) will eat it. BIS yu.

-Go, object pronoun (2s). You. Na-m ṣita-go I see you (singular). BIS yu.

hav ~ -av, negation marker (first morpheme). Na-n hav gani-a tehe I did not eat it. Go-s-av gani-a tehe go-v mate If you don’t eat it you will die.

-i¹, transitivising transformative. Transforms an intransitive verb to a transitive one. Na-m ate I look. Nam ate-i-ni-ra I look for them.

-i², see vi.

-m, see mwa.

-m-, see mwa.

men <=> -men, tense/apsect marker. Desiderative action/condition. Men hae He/she/it should go up. Go-men lai-a You (singular) should take it. See also mwa, nu, si, vi.

-men, see men.

mom, durational modifier. Still/yet. Ta-v mom gita-e We (inclusive) will still/yet see it.

-mu-, see mwa.

mwa <=> -m, -m- and -mu-, with -m also ~ -zero, tense/apsect marker. Present action or condition, action or condition in progress. Mwa maturu He sleeps/is sleeping. Na-m hae I go up. Ra-m-dol maturu They (three) sleep/are sleeping. Ra-mu-ru maturu They (two) sleep/are sleeping. Ra ban They go/are going. See also men, nu, si, vi.

-n. See nu.

na-, subject pronoun (1s). I. Na-m hae. I go up/am going up. BIS mi.

-na. See -ana.

-ni- <=> -zero-, indirect object ligature. Links verb to suffixed indirect object pronoun. Na-m ate I look. Nam ate-i-ni-ra I look for them. Go-m ate-i-au You are looking for me.

nu <=> -n and zero, tense/apsect marker. Past action or condition, completed action or condition. Nu mate He/she/it died. Ra-n mate They died. Ra-ru mate They (two) died. See also men, mwa, si, vi.

ra-, subject pronoun (3p). They (more than three people, or ref. two or three people in a context in which this has already been specified). Ra-m avo They speak/are speaking. BIS olgeta.

-ru, number marker (DU). Dual (two actors). Gi-ru vi vano You (two) will go.
-s, see si.

si <=> -s, tense/aspect marker. Conditional action or condition. Ra-s van If they go;
si gani-a If he/she eats it; gi-ru si gani-a If you (two) eat it. See also men, mwa,
nu, vi.

siv, durational modifier. Just. Ta-v siv gita-e We (inclusive) will just see it.

ta-, subject pronoun (1p.incl.). We (more than three people including addressee/
addressees, or two or three people including addressee/addressees in a context in
which this has already been specified). Ta-v hivo We (inclusive) will go down. Ta-ru vi hivo We (two inclusive) will go down. Ta-m-dol hivo We (three
inclusive) are going down. BIS yumi.

te, see tehe.

tehe ~ te, negation marker (second morpheme). Not. Go-n hav gita-e tehe You
(singular) did not see it.

-tol ~ -dol, number marker (TRL). Trial (three actors). Ra-tol vi gani-a They three
will eat it. Gi-m-dol g# gani-a You three are eating it.

-u, see -au.

-v, see vi.

va- <=> ba-, causative transformative. Cause to be. Vi va-mate-a He/she will kill
(lit. ‘cause to be dead’) him/her/it. Ra ba-mate-a They are killing (lit. ‘causing to
be dead’) him/her/it.

vi <=> -v, with -v also ~ -i², tense/aspect marker. Future action or condition, action
or condition not yet in progress. Vi hae He/she/it will go up. Ra-v hae They will
go up. Na-i vano I will go. Gi-ru vi hae You two will go up. See also men, mwa,
nu, si.

vi- <=> bi-, reciprocal transformative, compatible only with reduplicated verbs.
Indicates reciprocity of action or condition. Ra-n vi-ahiahi-ra They were kicking
their lower legs backwards reciprocally (referring to dancers doing different
dances in different groups concurrently). Ra bi-ahiahi They are kicking their
lower legs backwards reciprocally etc.

4 Coverage of verb phrase functors in the reference grammar

Three kinds of information about verb phrase functors need to be provided for the
dictionary user in the grammar portion of the Introduction:

(i) description of the phonological conditioning of allomorphs, including
consideration of any zero stratagems;

(ii) description of the morphological conditioning of allomorphs, again
including consideration of any zero stratagems;

(iii) an indication of the combinatorial possibilities of functors within the
preposed and postposed nuclear and peripheral slots.
This information, appropriately numbered to facilitate reference to it in the relevant
dictionary entries, will form part of the section in the Introduction dealing with the verb
phrase.

In the interests of providing an objective and precise baseline in the context of this
paper the idiom used in the following descriptions is that of the meta-language of
linguistics. However, as the dictionary Introduction is aimed at a general readership rather
than a linguistically expert one, the idiom of the eventual final English, Raga and Bislama
text will need to be primarily an everyday one, with informal exemplification tailored to
the discourse needs of each language. Given the complexity of some of the material this
will at times be no easy task.

All citations of dictionary entry head-words and of their allomorphs in this section are
in the bold type that the dictionary user will have experienced for such material in the
lexical entries.

4.1 Phonological conditioning

The conditioning specific to each of the phonologically conditioned allomorph sets
marked by ~ in the lexicographic presentation of verb phrase functors is detailed below.
Relevant examples are provided in §3.1.

- a ~ -e, with -a occurring after verb-final e, i, or u, and -e occurring after verb-final
  a or o.
- ana ~ -na, with -ana occurring after verb-final e, i, o, and u, and -na occurring
  after verb-final a.
- au ~ -u, with -au occurring after verb-final e, i, o, and u, and -u occurring after
  verb-final a.
- hav ~ -av, with -av occurring after -s, and hav occurring in all other possible
  contexts.
- m ~ zero, with zero occurring before verb-initial b except after na-, and -m
  occurring in all other possible contexts.
- tol ~ -dol, with -dol occurring after -m-, and -tol occurring in all other possible
  contexts.
- v ~ -i, with -i occurring before verb-initial v-, and -v occurring in all other
  possible contexts.

4.2 Morphological conditioning

The conditioning specific to each of the allomorph sets marked by <=> in the
lexicographic presentation of verb phrase functors is detailed below. Relevant examples
are provided in §3.1.

ge- <=> zero, with ge- occurring after bere ‘lest’, and zero occurring in all other
possible contexts.

---

5 The emic phonological rationale for the zero allomorph here is essentially an aversion to m before b
except in contexts consistent with the phonological constraint noted in §1.1 above, namely that /b/ has a
homorganically prenasalised allophone which is obligatory after NV(V) where N = nasal consonant and V
= vowel. This aversion leads to, for example, na-m ban I go, but go ban you (sg.) go, and ra ban they go.
men <=> -men, with men occurring after zero indication of 3s subject pronoun, and -men occurring in all other possible contexts.

mwa <=> -m, -m-, -mu-, with mwa occurring after zero indication of 3s subject pronoun, -m occurring after an overt subject pronoun (except where -m ~ zero as noted in §4.1), -m- occurring only between a non-singular subject pronoun and -dol number marker (TRL), and -mu- occurring only between a non-singular subject pronoun and -ru number marker (DU).

-m- <=> zero, with zero occurring before -au, and -ni- occurring in all other possible contexts.

nu <=> -n <=> zero, with nu occurring after zero indication of 3s subject pronoun, -n occurring after an overt subject pronoun that is not qualified by a number marker, and zero occurring between an overt subject pronoun + number marker and a verb.

si <=> -s, with si occurring after zero indication of 3s subject pronoun, and between a subject pronoun + number marker and a verb, and -s occurring after an overt subject pronoun.

va- <=> ba-, with ba- occurring in a primary verb phrase after mwa or any of its allomorphs or after any verb aspect marker ± hav + mom, and in a serial verb phrase following a primary verb phrase containing mwa or any of its allomorphs; and with va- occurring in all other contexts (see Walsh 1982:236–237, 1995:815).

vi <=> -v, with vi occurring after zero indication of 3s subject pronoun, and between a subject pronoun + number marker and a verb, and -v occurring after an overt subject pronoun (except where -v ~ -i as noted in §4.1).

vi- <=> bi-, with bi- occurring in a primary verb phrase after mwa or any of its allomorphs or after any verb aspect marker ± hav + mom, and in a serial verb phrase following a primary verb phrase containing mwa or any of its allomorphs; and with vi- occurring in all other contexts (see Walsh 1982:236–237, 1995:815).

4.3 Free variation

As no phonological or morphological conditioning is apparent for tehe ~ te these allomorphs are presumed to be in free variation, with speed of utterance possibly being a relevant variable.

4.4 Combinatorial possibilities

The ways in which the Raga verb phrase grammatical functors can be combined are outlined in this section, with the proviso that further investigation of some of these possibilities is still in progress.

4.4.1 For functors affixed to the verb

For most of these functors ordering is not an issue, as they are affixed directly to the verb. However, ordering is an issue with -ni- <=> zero and members of the paradigmatic
set of object pronouns. After a compound verb of which the transformative -i\(^1\) is a component, or after some specifiable non-compound verbs ending in i, or after an adverb interpolated into the verb phrase immediately after the verb, the occurrence of -ni- <= zero is obligatory before a phonologically appropriate member of the object pronoun set, i.e. before -a, -au, -go or -ra.

4.4.2 For peripheral functors in an active mood primary verb phrase

The combinatorial issues for these functors concern only the preposed elements, as the sole postposed element is part of the discontinuous negation marker hav ~ -av .... tehe ~ te. Except for when a primary verb phrase consists of an unmodified verb base used imperatively, the phrase has to begin with a segment (comprised of one or two words) consisting of + subject marker ± number marker + tense/aspect marker (with the ordering of the latter two elements being variable).

Depending on what is semantically appropriate in a given context, this segment can consist of any one of the following:

(i) + one of the subject pronoun morphemes na-, go-, ta-, ga-, gi-, ra- + one of the tense/aspect marker allomorphs -m ~ zero, -n, -v ~ -i\(^2\), -s, -men;
(ii) + one of the non-singular subject pronoun forms ta-, ga-, gi-, ra- + either the tense/aspect marker allomorph -mu- + the number marker morpheme -ru or the tense/aspect marker allomorph -m- + the number marker allomorph -dol;
(iii) + one of the non-singular subject pronoun morphemes ta-, ga-, gi-, ra- + either the number marker morpheme -ru or the number marker allomorph -tol + either the zero allomorph of the tense/aspect marker morpheme nu <= zero or one of the tense/aspect marker allomorphs vi, si, men;
(iv) + either the subject pronoun allomorph ge- + one of the tense/aspect marker allomorphs -m, -n, -v ~ -i\(^2\), -s, -men, or the zero allomorph of the subject pronoun morpheme ge- <= zero + one of the tense/aspect marker allomorphs mwa, nu, vi, si, men.

After this obligatory segment the following can occur: ± hav ~ -av (which entails tehe ~ te immediately postposed to the verb) ± siv ± mom.

4.4.3 For peripheral functors in a serial verb phrase

The obligatory initial functor in a serial verb phrase is ba, which, in addition to its ligature function, signifies that all the preposed peripheral functors in the preceding primary phrase apply also to the verb in the serial phrase. Negation in a primary phrase precludes that phrase being followed by a serial phrase. In terms of present knowledge the peripheral functor possibilities in a serial verb phrase are + ba ± siv ± mom provided that a given + element is not present in the preceding primary phrase.

4.4.4 For peripheral functors in a passive mood primary verb phrase

The obligatory initial functor in a passive mood verb phrase can be any one of the tense/aspect marker allomorphs mwa, nu, vi, si, men. Whether or not there are ± possibilities in such a phrase for any of hav..... tehe ~ te, siv or mom is under investigation.
5 Coda

Some issues which arise in considering the lexicographic presentation of Raga verb phrase functors have been outlined and my preferred response to these issues has been illustrated by listing the proposed dictionary entries for the functors, together with an indication of the kinds of explanatory material related to these entries which will be included in the grammatical sketch that will form part of the Introduction to the dictionary.

This explanatory material has been presented here in terms of the meta-language of linguistics, but in the actual text of the Introduction this presentation will as far as possible be in everyday English, accompanied by Raga and Bislama versions, rather than translations, of the English material.

References

‘Eating’ and ‘drinking’ in Kalam

ANNA WIERZBICKA

1 Are ‘eating’ and ‘drinking’ universal human concepts?

Are ‘eating’ and ‘drinking’ universal human concepts? Pawley and Bulmer’s dictionary of Kalam (in press) makes it clear that they are not and that, for example, Kalam has no such concepts. This doesn’t mean that Kalam people cannot speak and think about what English speakers call ‘eating’ and ‘drinking’, but that they speak, and presumably think, about such activities in a different way: not in terms of ‘eating’ and ‘drinking’ but in terms of a concept which they link with the verb ŋb-, roughly, ‘consume’ (and its variants). Pawley and Bulmer’s Kalam dictionary assigns to this verb as many as eight different meanings or senses. The most important, and presumably the most common, among these senses is the first one, described and illustrated in the Dictionary as follows:

Consume, especially by mouth; thus: 1. Eat, drink. ŋbin agen, ŋuurt gs-ap. ‘When I try to eat it hurts.’ (lit. When I say ‘I’ll eat!’ pain keeps happening.) ŋg pagen, apek ŋbin. ‘When I suck the liquid it comes up (through the straw) and I drink it.’ Tap etp nbsay? ‘What are they eating/drinking?’

Although the Dictionary glosses the Kalam concept with two English verbs (‘Eat, drink’), to its credit, it doesn’t regard the two glosses as two distinct meanings, but rather, recognises the unitary character of the Kalam concept (‘consume’). As the last example in the entry above illustrates, a sentence like Tap etp nbsay? ‘What are they eating/drinking?’ is vague rather than ambiguous. It may seem ambiguous from the point of view of an English speaker but not from a Kalam point of view: clearly, Kalam speakers do not habitually think about ‘oral consumption’ in terms of two distinct conceptual categories (1. eat, 2. drink), as speakers of English (and other European languages) do, but rather, in terms of one, unitary category.

As we have seen, the Kalam dictionary offers, in addition to the two glosses ‘eat, drink’, a unitary one: ‘consume (especially by mouth)’. Why can’t we say then that ŋb-, which makes no distinction between ‘eating’ and ‘drinking’, means the same as the English word consume? The addition ‘especially by mouth’ is no doubt motivated by other, secondary, uses of ŋb-, for example, as in ‘drown in the ocean, lit. consume sea-water’, but why can’t we say that in its primary meaning ŋb- means the same as consume (in its literal sense, that is, ‘consume by mouth’)?

651
The intuition behind glossing \( \tilde{n}b \)- as ‘eat, drink’ (rather than simply ‘consume’) reflects no doubt the basic and colloquial status of this Kalam word, similar to the basic and colloquial status of *eat* and *drink* rather than to the technical and, as it were, disjunctive character of *consume*. The *Collins Cobuild English language dictionary* (1991) defines *consume*, sensibly though not accurately, as ‘eat or drink’. The definition is not quite right because it doesn’t work in substitutions. For example, the sentence ‘they consumed vast quantities of alcohol’ does not mean that ‘they ate or drank vast quantities of alcohol’. One could say, however, that it means something along the following lines (very roughly): ‘They introduced alcohol into their bodies, by doing something to some alcohol as people do something to some things when they want to eat or drink these things’.

This very rough approximation suggests that the meaning of *consume* is based on the meaning of *eat* and *drink* (as the meaning of *parent* is based on the meanings of *mother* and *father*). To eliminate the reference to *eat* and *drink* in the definition of *consume* we could try to replace it with a reference to the *mouth*, along the following lines: ‘to consume something is to introduce it into one’s body through the mouth’. This would not work however, because this version would apply to tobacco or marijuana as well as to food and drink, which is not compatible with the ordinary use of *consume*.

It is also interesting to note that the verb *consume* is not used to describe human activities in progress, as *eat* and *drink* are in English and \( \tilde{n}b \)- is in Kalam. For example, it would be odd to say ‘I saw her consuming something’ or ‘What are they consuming?’. *Consume* is an abstract verb which indicates the speaker’s lack of interest in the physical activity as such and focusses on the outcome. Clearly, \( \tilde{n}b \)- is not like that: it is a verb of physical activity, analogous, broadly speaking, to the English verbs *eat* and *drink* rather than to the English verb *consume*. It describes a kind of physical activity — from an English point of view, *two* kinds (*eat* and *drink*), but from a Kalam point of view, *one* kind.

Arguably, therefore, it makes more sense to describe the meaning of \( \tilde{n}b \)- as ‘eat/drink’ than as ‘consume’, but clearly, this cannot be the end of the story. From a Kalam perspective, \( \tilde{n}b \)- is a unitary conceptual category, not a disjunctive one (*eat or drink*). To capture this unitary meaning of \( \tilde{n}b \)- from a Kalam perspective we need to go beyond ‘translating’ it into conceptual categories of English (*eat*, *drink*) and try to understand what such a unitary category could possibly mean for people who do not have separate conceptual categories of *eat* and *drink*.

Lexicography is not a traditional cultural pursuit in New Guinea. Nonetheless, it might be useful at this point to try to imagine that we are in the shoes of a hypothetical Kalam lexicographer trying to explain the meaning of \( \tilde{n}b \)- in a large monolingual dictionary of Kalam. How could such a hypothetical Kalam lexicographer (who doesn’t speak English) approach this task?

To answer this question (in English!) we need to forget, for a short time, English as we know it, that is, English in all its culture-specific elaboration, and try to limit our conceptual kit to a small set of concepts which English and Kalam share.

2 Universal human concepts and their Kalam exponents

Drawing on the experience of NSM researchers who have worked with many diverse languages of the world, I would hypothesise that Kalam, like all other languages, has words for simple concepts like ‘do’ and ‘happen’, ‘someone’ and ‘something’, ‘body’ and ‘part’, ‘time’ and ‘after’, ‘want’ and ‘because’, ‘inside’ and ‘near’, and another fifty or so
concepts (listed in Table 1 below), which evidence suggests are universal (lexicalised in all languages) and which can also be regarded as ‘semantic primes’, that is, elementary concepts in terms of which thousands of other, more complex, concepts can be adequately defined and explained. (For discussion of the NSM approach to semantics, see Goddard (1998); Wierzbicka (1996); Goddard and Wierzbicka, eds (2002); for the first application of NSM approach to Kalam, see Pawley (1994); for extensive discussion of Kalam from an NSM perspective see Goddard (2001); see also Wierzbicka (1996:200–202); Goddard (2002:27–28).)

Table 1: Semantic primes — English version

<table>
<thead>
<tr>
<th>Substantives:</th>
<th>I, YOU, SOMEONE, SOMETHING/THING, PEOPLE, BODY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomy, Partonomy:</td>
<td>KIND, PART</td>
</tr>
<tr>
<td>Determiners:</td>
<td>THIS, THE SAME, OTHER/ELSE</td>
</tr>
<tr>
<td>Quantifiers:</td>
<td>ONE, TWO, MUCH/MANY, SOME, ALL</td>
</tr>
<tr>
<td>Evaluators:</td>
<td>GOOD, BAD</td>
</tr>
<tr>
<td>Descriptors:</td>
<td>BIG, SMALL</td>
</tr>
<tr>
<td>Mental predicates:</td>
<td>THINK, KNOW, WANT, FEEL, SEE, HEAR</td>
</tr>
<tr>
<td>Speech:</td>
<td>SAY, WORDS, TRUE</td>
</tr>
<tr>
<td>Action, events, movement, contact:</td>
<td>DO, HAPPEN, MOVE, TOUCH</td>
</tr>
<tr>
<td>Location, existence, possession, specification:</td>
<td>BE (SOMEWHERE), THERE IS, HAVE, BE (SOMEONE/SOMETHING)</td>
</tr>
<tr>
<td>Life and death:</td>
<td>LIVE, DIE</td>
</tr>
<tr>
<td>Time:</td>
<td>WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT</td>
</tr>
<tr>
<td>Space:</td>
<td>WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE</td>
</tr>
<tr>
<td>Logical concepts:</td>
<td>NOT, MAYBE, CAN, BECAUSE, IF</td>
</tr>
<tr>
<td>Augmentor, intensifier:</td>
<td>MORE, VERY</td>
</tr>
<tr>
<td>Similarity:</td>
<td>LIKE</td>
</tr>
</tbody>
</table>

Notes: Primes exist as the meanings of lexical units (not at the level of lexemes) • Exponents of primes may be words, bound morphemes, or phrasemes • They can be formally complex • They can have different morphosyntactic properties, including word-class, in different languages • They can have combinatorial variants (allolexes) • Each prime has well-specified syntactic (combinatorial) properties. • Two (or more) primes can share the same lexical exponent, with different syntactic properties.

I do not hypothesise that all the concepts listed in Table 1 have distinct lexical exponents in Kalam, but only that they have identifiable ones. Some concepts listed in Table 1 share their lexical exponents in Kalam (as they do in many other languages). For example, the stem $g$- is the lexical exponent of both HAPPEN and DO. This doesn’t mean, however, that the concepts HAPPEN and DO cannot be distinguished in Kalam. For example, in sentence (1) below $g$- means, unambiguously, ‘happen’, when in sentence (2) it means, unambiguously, ‘do’ (both examples and glosses based on Pawley 1994).
(1)  *Mñab nb ak ned wagn ak g g-ek*  
country such this first origin the happen 3s.happen  
*mñab Aytol-jl alym.*  
country Aytol-jl down.there  
'The place where this originally happened was down there at Aytol-jl.'  
(Pawley 1994:408)

(2)  *Pen g-pan g-pin.*  
reciprocally 2s.did 1s.did  
'I did the same as you.' (Pawley 1994:404)

As noted in Goddard’s (2002:27-28) discussion of these sentences, the distinctive thing about (2) is the use of the reciprocal particle *pen*, ‘in return’, whereas the distinctive thing about (1) is the subject *ak*, ‘this’. Some other sentences, however, such as (3) below, can be ambiguous between ‘happen’ and ‘do’:

(3)  *Tap etp g-p?*  
thing what happen/do-3s:PF  
‘What’s happened?’ OR: ‘What has he done?’ (Based on Pawley 1994:408)

As Goddard (2002) points out, the analytical question posed by sentences like (3) is whether it is necessary to regard them as ambiguous, and therefore to treat *g*- as polysemous, or whether it is possible to sustain a monosemy interpretation. To quote Goddard (changing the numbers in referring to particular examples):

One could perhaps claim that *g*- has a unitary meaning in (1) and (2) which is Kalam-specific and simply cannot be stated in English, and that the apparent differences (from an English perspective) are imposed by the lexico-grammatical contexts. But if *g*- has a unitary meaning, then (3) cannot be ambiguous after all — which seems unlikely in the light of Pawley’s description. One salient difference between a ‘happen’ interpretation and a ‘do’ interpretation concerns ‘aboutness’ (cf. Sasse 1987). A question like ‘What happened?’ is not ABOUT any particular person, whereas ‘What did he/she do?’ is about a specific person. To claim that *g*- has a unitary meaning in (3) is to claim that this difference does not exist in Kalam, and that a Kalam speaker would be completely indifferent to it. If, as I assume, this is not the case, an interesting question arises: How could a Kalam speaker identify the distinct meanings involved, given that both are expressed by the same lexical form? Presumably — by reference to other, unambiguous, sentences. The Kalam speaker can say: ‘it can mean the same as *g*- in example (1) [i.e. ‘happen’], or it can mean the same as *g*- in example (2) [i.e. ‘do’].

What matters most in the present context is that if a Kalam speaker wanted to explain in Kalam what *ñb*- means, they could readily draw on the concepts DO and HAPPEN (using two exponents identical in their form but different in their grammatical properties). For example, he or she could say the exact equivalents of the following English sentences:

She was doing something to something with her mouth for some time.

Because of this something was happening to this something at the same time.

Thus, we can assume that it would not be very difficult to draw a Kalam version of Table 1. This version would show a good deal of polysemy, for example, the same lexical element *g*- would be presented as the Kalam counterpart of the English elements DO and HAPPEN, but this by itself would not be a problem from the point of view of developing a
Kalam explication of complex Kalam concepts, for example, of the complex Kalam concept nd- (‘eat/drink’).

Furthermore, we can also hypothesise that Kalam, like other languages, has a number of ‘semantic molecules’, including, for example, ‘hands’, ‘head’, and ‘mouth’, which are definable in terms of the semantic primes but which can enter the meanings of many other more complex concepts as semantic chunks (Goddard 2007; Wierzbicka 2007, in press a).

3 The Kalam ‘eat/drink’-word seen from a Kalam perspective

Experimenting with various semantic explications and drawing on the semantic template developed for other verbs of physical activity (Goddard and Wierzbicka 2008), I have come to the conclusion that the hypothetical Kalam lexicographer could define the meaning of nd- as in the explication below. With one exception, this explication relies exclusively on universal semantic primes (including ‘do’, ‘happen’, ‘someone’, ‘something’, ‘people’, ‘body’). The exception is the word mouth, which can be plausibly regarded as a near-universal semantic molecule (not necessarily in the sense that words for ‘mouth’ would mean exactly the same in all languages, but that they would have a universal semantic core, identifiable through NSM).

The Kalam word nd- (‘eat/drink’)

Someone (X) was nd-ing something (Y)

LEXICO-SYNTACTIC FRAME
someone (X) was doing something to something (Y)
with their mouth[M] for some time
because of this, something was happening to this something (Y) at the same time

PROTOTYPICAL MOTIVATION
X was doing it like people do something to something
when they are doing something to this something with their mouth[M]
because they want part of this something to be inside their body

MANNER
when someone does something like this to something
the same thing happens many times
it happens like this:
this someone does something to something with their mouth[M]
after this, part of this something is for a short time inside this someone’s mouth[M]
after this, this someone does something to it with their mouth[M]
because of this, after this, it is not inside this someone’s mouth[M] anymore,
it is inside another part of this someone’s body for some time

POTENTIAL OUTCOME
if someone does this to something for some time
after some time all parts of this something can be inside this someone’s body

The template used in this explication was not invented ad hoc for the purposes of explicating nd- in Kalam and eat and drink in English, but was developed on the basis of extensive experimentation with many verbs of physical activity in many languages (see Goddard and Wierzbicka 2008; Wong, Goddard and Wierzbicka, to appear).

Briefly, the ‘lexico-syntactic frame’ accounts for a given verb’s basic syntactic profile, including in particular its transitivity and its temporal and aspectual properties. The ‘prototypical motivation’ sets up a prototypical situation, with a prototypical goal which shapes, to a large extent, the details of the ‘manner’ in which the action is done. The ‘manner’ segment specifies the constant features of the way in which the action is
performed. The final segment, the ‘potential outcome’, is closely related to the ‘prototypical motivation’: if the action proceeds in accordance with the ‘prototypical motivation’ for some time, the goal set out in the ‘prototypical motivation’ section can be achieved.

In the case of the Kalam verb ñb-, the ‘lexico-syntactic frame’ specifies that the action is performed by someone for some time and that it affects, simultaneously, some things or some substance. The ‘prototypical motivation’ refers to ‘doing something to something with one’s mouth’ in order to introduce part of this something into one’s mouth. The ‘manner’ shows that the action is iterative and it describes its stages: doing something to something with one’s mouth, having a bit of that thing in one’s mouth for a short time, doing something to it with one’s mouth (i.e. swallowing it), after which the bit in question travels to some other place in the person’s body (i.e. the stomach). The ‘potential outcome’ envisages the possibility of the whole thing in question ending up inside the person’s body.

4 The English verbs ‘eat’ and ‘drink’ seen from a Kalam perspective

Suppose now that our hypothetical Kalam lexicographer seeks to understand, and to explain to other Kalam speakers, the meaning of the two unfamiliar English concepts, with no equivalents in Kalam: ‘eat’ and ‘drink’. There can be little doubt that this lexicographer would try to explain the meaning of the English verbs *eat* and *drink* via the Kalam word ñb-. As a starting point, he/she might offer the explanation that *eat* and *drink* stand for two varieties of ñb-ing, and that the distinction between the two is based, partly, on the kind of thing that the ñb-ing person is putting inside his or her mouth. If this thing is like water, the English speaker would choose *drink*, and if it is not like water, they would choose *eat*.

Of course such an explanation will work in Kalam only if Kalam, too, has a word for ‘water’, and if this word is very close, if not identical, in meaning to the English word *water*, and Pawley and Bulmer’s dictionary tells us that it does. Since ‘water’ is not a universal semantic prime, the availability of a word for it in Kalam cannot be taken for granted. In English, ‘water’ is a semantic molecule, which enters — alongside the molecule ‘mouth’ — the meanings of the words *drink*, *eat*, and of course many others (including, for example, *liquid*, *cup*, *mug*, *bottle*, *jug*, *urine*, *tears* and many others). Presumably, the Kalam word for water, ñg, is also a semantic molecule in Kalam, but evidently this molecule is not included in the meaning of the verb ñb- (‘eat/drink’).

In addition to the distinction between ‘things like water’ and ‘things not like water’, an astute Kalam lexicographer would point out that the English verbs *eat* and *drink* refer also to a difference in what is happening in the mouth of the person ñb-ing (‘consuming’) something. If the substance ‘consumed’ is like water the following happens to it many times: first some of it is for a very short time in the person’s mouth, then the person does something to it with the mouth (‘swallows’ it), and then it is not in the person’s mouth any longer. If, on the other hand, the substance ‘consumed’ is *not* like water, what happens to it each time is more complex: when some of it is in the person’s mouth, the person does something to it with some parts of the mouth (‘chews’ it, ‘bites’ it, moves it around with the tongue), and as a result, something happens to it; after this, the person does something else to it with their mouth (‘swallows’ it), and after this, because of this, it is no longer in this person’s mouth.

---

1 In English, a more idiomatic rendering for ‘all parts of this something’ would be ‘all of it’.

2 The idea that not only drink but also eat may refer in its meaning to ‘things like water’ (in the case of eat, ‘things not like water’) is due to Michael Neubauer (pers. comm.).
If our hypothetical Kalam lexicographer persists in the analysis of the alien English concepts ‘eat’ and ‘drink’ they might notice that there is a third difference between them, related to the other two: ‘drinking’ can be done directly with one’s mouth (for example, when someone is ‘drinking’ water from a source), whereas ‘eating’ typically involves also the use of the hands.

On the basis of all these observations the hypothetical Kalam lexicographer could arrive at the following explication of the English words eat and drink, stated here in NSM English but statable also in NSM Kalam, by means of the Kalam set of semantic primes and the Kalam semantic molecules ‘mouth’ and ‘water’:

**Someone (X) was eating something (Y)**

Someone (X) was doing something to something (Y)  
because of this, something was happening to this thing (Y) at the same time

X was doing it like people do something not like water  
when they are doing something to this something with their mouth[M]  
because they want part of this something to be inside their body

when someone does something like this to something
the same thing happens many times
it happens like this:
this someone does something to something with their hands[M]
after this, part of this thing is for a short time inside this someone’s mouth[M]
when it is inside this someone’s mouth[M],
this someone does something to it with some parts of their mouth[M]  
because of this, something happens to it at this time
after this, this someone does something else to it with their mouth[M]  
because of this, after this, it is not inside this someone’s mouth[M] anymore,  
it is in another part of this someone’s body

if someone does this to something for some time
after some time all parts of this something can be inside this someone’s body

This explication of eat overlaps to a large extent with that of the Kalam verb ſb-, but there are also some differences. Thus, the ‘prototypical motivation’ of eat specifies that the agent is doing something with ‘something not like water’. This specification is missing in the case of ſb-. Its presence in the ‘prototypical motivation’ of eat does not imply that eat can only be applied to solids and that one cannot, for example, eat soup. It only suggests that the prototypical situation of eating involves solids.

The ‘manner’ of eating includes a reference to ‘the hands’, and also, to something happening to the bit in one’s mouth when it is there (e.g. being chewed). Both these references are absent from the explication of ſb- and both require a comment.

As for the use of the hands, I should clarify that I regard the verb eat as polysemous and that the meaning explicated here applies only to human eating. When animals or birds ‘eat’, they ‘eat’. I would claim, by analogy with human ‘eating’. Just as the ‘head’ of a snake is not the uppermost part of the snake’s body (as the human ‘head’ is) but rather, that part of the snake’s body which is ‘like’ the head in people’s bodies, so the snake’s ‘eating’, too, is conceived by analogy with human eating. (For discussion, see Wierzbicka 2007:37–38; 1980:86–88.)
The conceptual salience of the ‘hands’ in eat is of course related to the fact that prototypically, people eat solids, and that in general, when they eat something, they do so in the way in which people eat solids. This conceptual salience of solids (‘something not like water’) is also linked with the fact that when someone eats something this someone does something to the bits in his or her mouth (e.g., chews them) and that consequently, something happens to these bits while they are in this person’s mouth (they can be broken into smaller bits, mixed with saliva, and so on). These two components are also missing from the explication of ēb-.

As for the explication of drink, it is much closer to that of ēb-, but unlike the explication of ēb- it specifies that the substance introduced into one’s body is ‘something like water’.

**Someone (X) was drinking something (Y)**

someone (X) was doing something to something (Y) with their mouth[M]  
for some time  
because of this, something was happening to this something at the same time  

X was doing it like people do something to something like water[M]  
when they are doing something to this something with their mouth[M]  
because they want part of this something to be inside their body  

when someone does something like this to something  
the same thing happens many times  
it happens like this:  
this someone does something to something like water with their mouth[M]  
after this, because of this, part of this something is for a very short time  
inside this someone’s mouth[M]  
after this, this someone does something to it  
with their mouth[M]  
because of this, after this, it is not inside this someone’s mouth[M] anymore  
it is somewhere else inside this someone’s body  

if someone does this to something for some time  
after some time all parts of this something can be inside this someone’s body  

The explications of eat and drink developed here can of course be arrived at from within English: they don’t require an outsider’s, for example a Kalam person’s, perspective. However, to take a Kalam perspective helps us to de-naturalise English and to resist the temptation to interpret the indigenous Kalam concept ēb- through the prism of the English concepts ‘eat’ and ‘drink’. Eating and drinking are universal human activities, but so is ēb-ing. Universally, people put various substances into their mouth and swallow them, bit by bit. The process of doing so can take different forms and can involve different kinds of substances. How the activities involved are conceptualised and categorised depends, to some extent, on the language. The use of a mini-language based on universal human concepts (NSM) allows us to study those different conceptualisations and categorisations from a neutral, language-independent perspective. It frees us from our conceptual dependence on languages like English and at the same time it allows us to see these languages in a fresh way. Seen from a Kalam perspective, English is an exotic language, just as Kalam is exotic from the perspective of native speakers of English.
5 Words, concepts, and cultural practices

Why does English distinguish lexically between ‘eating’ and ‘drinking’ whereas, for example, Kalam does not? Presumably, culture-specific conceptualisation of human activities is related to salient culture-specific practices. One possible factor which could play a role here (suggested by Carol Priestly, pers. comm.) is the important cultural practice of ‘sucking’ sugar cane in Papua, ritualised in the custom of offering sugar cane as a refreshment, to visitors. Since one neither ‘eats’ nor ‘drinks’ sugar cane this practice may blur the distinction which seems so obvious to speakers of languages like English. There may, of course, be other contributing cultural factors.3

It should also be noted that English, too, has its own culture-specific conceptual categories in the area under discussion. One of them is the nominal category of ‘drinks’ – prototypically, alcoholic and social (‘consumed’ at particular times, in the company of other people). The verb drink, too, has a distinct meaning associated with alcohol. The Collins Cobuild English language dictionary (1991) defines this meaning as follows: ‘to drink means to drink alcohol, especially regularly or in large quantities, e.g. She said she didn’t smoke or drink ... You shouldn’t drink and drive ... I never drink alone.’

Thus, the conceptual categories ‘eat’ and ‘drink’ do not cut nature at its joints. The conceptualisation of such activities depends, to some extent, on what is culturally salient in a particular society. The concepts of DOING and HAPPENING are universal, as are also PEOPLE, BODY, INSIDE, and the other sixty or so concepts specified in Table 1. However, ‘eat’ and ‘drink’, while very common, are not universal.

Detailed lexicographic studies of non-European and not easily accessible languages such as Kalam can help speakers of English, including Anglophone scholars, to see the world in a new, non-Eurocentric, way. There is an important lesson here for colour studies, emotion studies, and many other areas of psychology and related disciplines.

The more science becomes dominated by English the more it needs to seek what Russian formalists of the early twentieth century used to call ostranenie (the principle of ‘making strange’, that is, of making the familiar look foreign). Detailed lexicographic studies of languages like Kalam can contribute significantly to that ostranenie, which the globalised and English-dependent contemporary science badly needs. They can teach us — if anything can — that apart from the five or six dozen empirically discovered universal semantic primes no other concepts can be taken for granted as culture-independent analytical tools — not even as seemingly simple and natural ones as ‘eat’ and ‘drink’.

References


3 For examples of other languages which do not distinguish lexically between ‘eating’ and ‘drinking’ see Wierzbicka (in press b.). Note that the explications of ‘eat’ and ‘drink’ give in that paper represent slightly revised versions of those included here.


—— in press a, The theory of the mental lexicon. In Tilman Berger, Karl Gutschmidt, Sebastian Kempgen and Peter Kosta, eds Die slavischen Sprachen (Handbücher zur Sprach- und Kommunikationswissenschaft (HSK) [the Slavic languages (Handbook of linguistics and communication science)]. Berlin: Mouton de Gruyter.


~ Epilogue ~
Taking up the ‘Pawley challenge’: speech formulae and linguistic theory

NIKOLAUS P. HIMMELMANN

0 Introduction

As briefly explained in the preface, the contributions to this festschrift had to be severely limited in a number of ways in order to keep it within manageable bounds. In line with these limitations, all chapters are primarily concerned with Austronesian or Papuan languages and cultures. But this book would be seriously incomplete as a celebration of Andrew Pawley’s achievements in linguistics in particular, if it did not contain an explicit, though minimal acknowledgement of his contribution to contemporary linguistic theory.

The core of his contribution consists in the following challenge: structural linguistics, broadly conceived, is incapable of handling two interrelated core components of linguistic competence, that is native-like selection (idiomaticity) and native-like fluency. This failure is most obvious in its inability to deal with the formulaic nature of all kinds of linguistic interactions, which are replete with speech formulae of various levels of abstractness (collocations, idioms, patterns for recounting events, etc.). Structural linguistics has failed to develop an adequate descriptive practice for speech formulae and lacks a convincing theoretical account of their structural properties and central role in human language. Such an account, among other things, would have to be psychologically realistic in the sense of being based on a ‘profound understanding of how human minds actually organize information’ (Pawley 1993:105) as exemplified by the one-clause-at-a-time hypothesis proposed by Pawley and Syder (2000).

In the 1970s, when first articulating the challenge, speech formulae in fact tended to be considered marginal phenomena not at all relevant to linguistic theorising. At the beginning, then, part of the challenge consisted in getting their central role in human language acknowledged within mainstream linguistics. In this regard, there have been

---

1 I am grateful to René Schiering, Sonja Riesberg, Malcolm Ross and Edgar Suter for helpful comments on an earlier draft of this homage.

2 That is, including not only classical structuralism, especially in its American descriptive guise, but all varieties of generative grammar and its post-generative successors and opponents. Note, however, that most European linguistic structuralisms and their offspring have differed here in at least one important regard which we will turn to in §3.
major changes in the last two decades, so that currently their importance for linguistic theory is no longer in doubt (cf. Pawley 2001b, 2007).

The deeper reason for the failure of structuralist linguistics to successfully deal with speech formulae, Pawley argues, is to be found in its reductionist conception of language. This conception of language, variously called grammarians’ language or the (parsimonious) grammar-lexicon model by Pawley, is based on the idea that the linguistically relevant aspect of language is that of a self-contained system of atomic units and combinatorial rules for these units and that the task of linguists consists in capturing this system in as economical and elegant terms as possible. What is ignored here almost completely is the fact that linguistic expressions on all levels of complexity — ranging from simple words to complex sentences — are social institutions, and that in order to arrive at a reasonably coherent and comprehensive view of how languages work, descriptive practice as well as linguistic theories have to incorporate methodologies, theoretical components, and representations for social constraints on linguistic units and knowledge.

The central role of speech formulae in natural language is the paradigm example of such social constraints: while the grammatical rules of a language — as conceived of in structuralist theory — may provide for numerous ways of expressing a given idea or state of affairs, there is usually only a single conventional (= native-like, fluent, idiomatic) way to express it. Hence, to repeat one of Pawley’s most often used examples, while there may be many grammatical ways of telling the time in (standard) English (e.g. three quarters before one, twelve and fifteen minutes, eleven plus two hours minus forty five minutes) there is only one native-like way for doing so, using the established formula X minutes/a quarter past/to HOUR (hence a quarter past twelve in this example).

In all his writings on the topic, Pawley emphasises the fact that other conceptualisations of language, which recognise the importance of speech formulae and hence address the issue of languages as social institutions, are not only possible but actually attested. Such alternative approaches — labelled humanists’ language, subject matter-model or vernacular view of language by him — can be found in disciplines and professions such as (traditional/commercial) lexicography, anthropology (ritual speech, performance routines), sociology (conversation analysis), language teaching, and ordinary language philosophy. Consequently, the challenge is to further develop structuralist theory and practice so that it becomes compatible with, and incorporates insights and practices from, the subject matter-models of language developed in these disciplines.

While Pawley was one of the first linguists in the 1970s to point out the limitations of the parsimonious grammar-lexicon model and to emphasise the need for an appropriate place for speech formulae in linguistic theory, other linguists have raised similar concerns, in particular in the late 1970s and 1980s. As a consequence, linguistics now looks quite different compared to the 1960s when Chomsky’s Aspects Model was the most advanced and articulated instantiation of the parsimonious grammar-lexicon model and the major target for critics of all persuasions. Thus, the question arises whether and to what extent structural linguistics has taken up the ‘Pawley challenge’. This epilogue attempts a very brief and highly tentative first assessment.

One way in which the Pawley challenge differs from similar critiques of the structuralist program pertains to the fact that one of its two major motivations was a deep lack of satisfaction with structuralist descriptive practice regarding little-known languages,

---

3 The other major motivation was the joint work with Syder on the syntax and discourse structure of conversational speech as reported in Pawley and Syder (1983a, 1983b, 2000).
resulting from his own work on Kalam (Pawley 1993 [first presented in 1980]). Although theory and practice are never fully and neatly separable, it will be useful to single out the question of whether and to what extent descriptive practice regarding little-known languages has changed since Pawley first presented his challenge in 1980. Section 1 will be concerned with this question, while the more comprehensive question regarding changes in the basic conceptualisation of language as the object of linguistic study is taken up in §2 and §3. Section 2 deals with current usage-based approaches, and §3 with the distinction between system and norm, which has been used in quite a number of European structuralist frameworks.

1 Language documentation avant la lettre

The question of whether descriptive practice regarding little-known languages has changed significantly since the 1960s when Pawley did his main descriptive work on Kalam has a straight yes/no answer. That answer is ‘no’ in that to date no real progress has been made in developing a descriptive practice which allocates a central role to speech formulae and is capable of representing and giving an analytic account of the knowledge underlying native-like selection and fluency.4

But the answer is ‘yes’ in that traditional descriptive practice has been expanded by a new subfield called documentary linguistics which dissociates the agenda for collecting and processing linguistic data from a narrowly structuralist view of language in terms of the grammar-lexicon model. This view is replaced by a view where — in line with the Pawley challenge — the observable linguistic behaviour and native speakers’ metalinguistic knowledge are the major concern. Unsurprisingly, Pawley’s writings have been a major source of inspiration in arguing for and developing documentary linguistics (cf. for example, Himmelmann 1996:327–329, 1998:164, fn.6).

Note that, strictly speaking, documentary linguistics does not solve the speech formula problem, because it does not provide an analytic-representational format for them. Nevertheless, because they take into account both grammarians’ and humanists’ views of language, data collections compiled according to the documentary agenda will, ideally, contain many and variegated specimens of formulaic speech and hence furnish the empirical basis for an in-depth study of the structure and use of speech formulae in the documented community.

Pawley’s contribution to documentary linguistics, however, is not restricted to providing a crucial argument for its theoretical foundations. He has also been influential with regard to many aspects of the documentary agenda through his exemplary fieldwork practices, developed in collaboration with his teacher and colleague Ralph Bulmer. Of major import here are the following points where his and Bulmer’s practices often diverge significantly from those of their contemporaries:

4 N.B. this is not to say that no significant changes have happened in descriptive grammar writing since the 1960s. On the contrary, during the 1970s and 1980s the typological-descriptive framework for writing grammars of little-known languages has emerged and, in many respects, constitutes a significant advance over older descriptive practice (see the contributions in Ameka et al. 2006 for a recent assessment). However, these changes do not, or only very indirectly, address the Pawley challenge.
interdisciplinarity: Pawley has done all his fieldwork in interdisciplinary partnerships and repeatedly emphasised the need for interdisciplinary co-operation in working on and with little-known speech communities (e.g. Pawley 2001a).

long-term commitment: Pawley has stayed in contact with the speakers and communities he has worked with over many decades, keeps them abreast of his work and helps out as much as possible in their regularly recurring economic and health crises.

native speaker involvement: Pawley has trained and encouraged native speakers to become active partners in shaping and carrying out the research agendas concerning their communities. Among other things, this has meant that native speakers have been co-authors of many articles and books (e.g. Majnep and Bulmer 1977, 1990; Pawley and Sayaba 1971, 1982, 2003; Pawley, Gi, Kias and Majnep 2000; Majnep and Pawley 2001), that they have stayed and worked with Pawley for extended periods of time in New Zealand and Australia, and that they have accompanied him to international conferences and meetings.

In short, Pawley is a documentary linguist *avant la lettre* who has implemented the documentary agenda since his very first fieldwork with the exception that the massive amount of primary data he has compiled to date have not yet been publicly archived in such a way that they can be of use to future generations. However, he has spent a significant part of his professional life in preparing two voluminous dictionaries for his two major field languages, Kalam (Pawley and Bulmer 2003) and Waya Fijian (Pawley and Sayaba 2003), which contain substantial amounts of this data in an edited format.

2 Changes in (mainstream) linguistic theory

Has linguistic theory, and more specifically, grammatical theory responded to the Pawley challenge? At first, the answer would appear to be an unequivocal yes, because there are quite a few more recent theoretical frameworks and approaches which address many of Pawley's concerns (usually without explicitly referring to his work). These approaches have been referred to as *usage-based* and would include — in addition to much so-called functionalist work — *Construction Grammar* in all of its different guises. Among the concerns addressed in this work are the following ones:

- Many usage-based approaches, in particular Construction Grammar, place speech formulae at the centre of the theory and claim that all other lexically and grammatically relevant features of natural languages can be captured with the apparatus designed to give a full account of speech formulae.

- For most usage-based approaches, economy and parsimony in representation and conceptual apparatus are not a major concern. That is, they are happy to allow multiple, at least partially overlapping representations for the same unit. This is particularly obvious in *Exemplar Theory* (Pierrehumbert 2001, Bybee 2006) which assumes that speakers store all tokens of all the linguistic units they hear or produce, at least for some time (tokens not reactivated and reinforced by similar tokens decay over time).

- Most usage-based approaches attempt to be psychologically realistic by incorporating insights from psychology into how the human mind works.
• Usage-based approaches allow for gradience and differences in productivity and frequency on all levels of linguistic structure.

Despite this impressive (and far from comprehensive) list of changes — their proponents would say ‘improvements’ — in linguistic theory over the last three decades, however, it is debatable whether these changes suffice to meet the Pawley challenge. Here are a few points indicating that they, in fact, fall short of the still more radical changes called for by this challenge.²

To begin with, the key concepts of native-like selection and fluency do not play any role in usage-based approaches, neither explicitly nor implicitly. That is, usage-based approaches do not even remotely approach the type of comprehensive model of idiomatic competence that Pawley is envisioning. More importantly, it is not clear whether they are actually striving for such a model. While usage-based approaches differ significantly in the articulation of their theoretical goals and underpinnings and in their deviation from, or revision of, the structuralist agenda, they all tend to stick to some of its, apparently deeply entrenched, methods and ideas.

Thus, for example, usage-based theories and models do not concern themselves with questions of comprehensiveness regarding data coverage. Instead, they continue the Chomskyan ‘tradition’ of basing far-reaching claims on the analysis of a comparatively minute set of phenomena. Obviously, comprehensive accounts of how a natural language works are no small achievement and require the work of many people for many decades. Hence the point here is not that Construction Grammar or Exemplar Theory to date have not yet produced anything remotely approaching a comprehensive framework for analysing natural languages. The point is whether CG or ET are at all capable of and interested in producing something coming close to such a comprehensive framework.

Usage-based approaches also follow the highly problematic Chomskyan strategy of leaving the precise boundaries of its object of study undefined. Of course, Chomskyan theory targets U(niversal) G(rammar), understood as the genetically inherited code which defines the universal structural core of human language. In this sense, the object of study is clearly defined. But exactly which structural features attested in natural languages manifest or realise this structural core is unclear. More importantly, it is not at all clear that the theory is, at least in principle, capable of providing an unequivocal delimitation of the relevant structural phenomena. Even though usage-based approaches generally do away with the distinction between theoretically ‘interesting’ core structures and ‘uninteresting’ peripheral phenomena — which is constitutive of the Chomskyan approach — they still have a similar problem in that they also have to indicate in some way what they consider to be ‘the limits of language’ and thereby delimit the object their theories and models are targeted at.

And, looking at the types of data actually investigated in these approaches, we may note that indeed, there is only one new data type that has been added to the empirical agenda, that is the type of speech formula represented by partially variable phrase- and clause-sized patterns such as What’s X doing Y (Kay and Fillmore 1999). Otherwise, however, usage-based approaches largely concern themselves with the same phenomena as have been on the structuralist agenda for most of the 20th century. Thus, to date, they really have not yet started the vast expansion of the empirical agenda for linguistics implied in the Pawley challenge.

² In Pawley (2008:§5), he himself points out a few problems and desiderata which will not be repeated here.
Nevertheless, there is no doubt that in some regards usage-based approaches diverge significantly from major tenets of the structuralist agenda, the most important one perhaps being the relaxation, if not total abolition, of the principles of economy and parsimony in representation and conceptual apparatus mentioned above. But this move is also not without its problems because, as a consequence, usage-based approaches tend to neglect, or even explicitly deny, the relevance of structural generalisations, that is genuinely linguistic, as opposed to more broadly functional or cognitive principles or laws of language. Specifically, they tend to neglect the grammar part of the grammar-lexicon model, that is the rules for productively forming new expressions, and greatly expand the list part of this model.\(^6\)

But even if a large part of everyday linguistic behaviour is automatised and consists in replicating more or less fixed expressions of various degrees of complexity, there is also ample evidence that speakers can analyse these expressions and create new ones. Consequently, the Pawley challenge calls for theories and models which account for the fact that native speakers know, and are capable of doing, two things.

(a) they know and use large numbers of at least partially prefabricated linguistic expressions of various sizes, where ‘knowing’ importantly includes not only meaning and form of the expression but also its social ‘value’ (when is it appropriate to use the expression, thereby achieving which interactional goal?).

(b) they are able to take linguistic expressions of any size apart and create new ones from their parts in such a way that the meaning and intention conveyed by the new expression is understood by their interlocutors.

Hence, a comprehensive and realistic model of language has to contain components for both holistic and analytic (‘generative’) processing of linguistic expressions, as argued by, for example, Wray (2002) for ‘the lexicon’ and Jacobs (2008) for ‘constructions’. Wray (and many others thinking along these lines) remains silent with regard to what exactly the analytic component is supposed to look like and hence how the two components interact. But this is what I take the Pawley challenge to be in its most precise and ambitious reading: to develop the structuralist model of language in such a way that it can be synthesised with concepts and practices from subject matter approaches, thus accounting for linguistic competence both in its highly creative-analytic as well as in its thoroughly entrenched and socially constrained aspects.\(^7\) A core issue for this enterprise is the question of whether the machinery developed as part of the grammar-lexicon model to account for productive linguistic abilities and knowledge needs a basic and radical overhaul for such a synthesis to become possible, or whether some minor modifications will do.

---

\(^6\) As noted in Pawley (2008), the expansion of the lexicon in usage-based approaches tends to be done with the conceptual tools and machinery of the grammar-lexicon model, hence not taking on board the concepts and practices characteristic of subject matter models and thus not paying due attention to the social constraints on the listed expressions and patterns.

\(^7\) This view can also be found among some proponents of usage-based approaches, for example, Pierrehumbert who writes (2001:139): ‘The real challenge arises from the fact that the classical view does provide important insights about the mental representation of phonology. Although a word may have idiosyncratic phonetic properties, it is perceived as made up of units of sound structure which are also shared with other words. […] Thus, the correct model must describe the interaction of word-specific phonetic detail with more general principles of phonological structure.’
In concluding these observations on the uptake of the Pawley challenge, it may be useful to note that the basic problem underlying the challenge has been tackled in European linguistics for quite some time, although not really successfully, but in a way that differs in an interesting way from Pawley’s approach to it. The key for providing a place for idiomatic competence in a comprehensive theory of language in the European tradition is the distinction between system and norm, which has been intensely debated in European linguistics for much of the 20th century.8 The language system is conceived of as an abstract system of oppositions and rules which define what a possible linguistic expression in a given language is. The less abstract level of norm pertains to which of these merely possible expressions are actually in common use. Importantly, this notion of norm (or usage) is a descriptive one. It does not pertain to what some authority considers ‘good usage’ — this would be a prescriptive norm. Instead, it pertains to what is conventional in a given speech community and hence characteristic of the normal linguistic behaviour of its members.

The Pawley challenge and this system/norm distinction differ in many ways due to a fundamental difference in perspective and motivation. The former is empirically motivated and results from work with naturalistic (specifically conversational) data and fieldwork experiences. The European distinction is rooted in the notorious ambiguities of Saussure’s distinction between langue and parole and was developed as a purely theoretical (philosophical) argument (cf. Coseriu 1952/75) with no concern at all for descriptive practice.9

A very conspicuous difference between the two approaches is the fact that Coseriu’s prime examples for the distinction are phonological ones (Coseriu 1952/75:64–68):10 phonemes are abstract system units, while (allo-)phones — as conceived of in structuralist accounts — indicate the normal realisations of a given phoneme. Thus, for example, the phoneme /o/ in a 5-vowel system /i e a o u/ in principle allows a number of more open or more closed realisations. But in a given speech community, there will only be a single normal realisation for this phoneme — or a small number of positionally determined

---

8 See Coseriu (1952/1975) for the perhaps clearest and most convincing exposition of this distinction. This essay also includes a detailed review of related distinctions proposed in the first part of the 20th century in work by Jespersen, Hjelmslev, Brøndal, Bally, Sechehaye, Martinet, Gardiner, Trubetzkoy, Bühler, Porzig and several others.
9 This does not mean that the theoretical argument has not had any practical consequences. The fact that phraseology has been a major concern in various schools of European linguistics is, at least in part, motivated by this argument.
10 Coseriu also provides examples pertaining to other levels of linguistic structure. For inflectional morphology, the difference is illustrated with examples where the system allows two options only one of which is commonly used (e.g. the plural of Rumanian chibrit ‘match’ is either chibrite or chibrituri, but the later is the more normal/widely used one (1952/75:69); compare also English formulas versus formulae). Interestingly, he also includes the example of overgeneralisations found in children’s speech. Thus, for instance, oxes (instead of oxen) is said be a systematically possible but not normally realised plural of English ox. For derivational morphology, Coseriu refers to the well-known distinction between a merely systematically possible and a commonly used and hence traditional formation (compare actor/actress versus doctor/?*doctress). For the syntactic level, he gives two types of examples (1952/75:76): on the one hand, speech formulae and on the other hand, ‘normal/unmarked’ order in ‘free’ word order languages such as Latin. A major problem arises with regard to the lexical level (1952/75:77–80) because here it is difficult to make a distinction between lexemes as system units as opposed to lexemes as normative units.
variants — which define native-like speech behaviour. Similarly, in most languages there
is no contrast between alveolar and uvular trills as realisations of an /r/-phoneme and hence
both are possible realisations, but usually only one is the normal or traditional one.

With regard to phonology, then, it would appear that structuralist linguistics has — right
from its beginnings — provided a framework for both system and norm, both the general
structure underlying productivity and the specifics characterising idiomatic competence.
This is further corroborated by the fact that it has been possible — unproblematically! —
to link the basic concepts of this framework with social variables and hence to account for
social constraints on the linguistic system (cf. variationist sociolinguistics, especially of the
Labovian style). If this is indeed the case, the Pawley challenge does not arise for
phonology. And it may be a seemingly circumstantial, but perhaps nevertheless productive
new response to the Pawley challenge to further explore the question of why this is so and
what, possibly, can be learned from phonology in order to meet the challenge on other
levels, in particular derivational morphology and syntax (speech formulae).

References

Ameka, Felix, Alan Dench and Nicholas Evans, eds, 2006, Catching language:
the standing challenge of grammar writing. Berlin: Mouton de Gruyter.

Bybee, Joan L., 2006, From usage to grammar: the mind’s response to repetition.
Language 82:711–733.

Bybee, Joan L. and Paul Hopper, eds, 2001, Frequency and the emergence of linguistic
structure. Amsterdam: Benjamins.

Coseriu, Eugenio, 1952/75, System, Norm und Rede. In Sprachtheorie und Allgemeine
Sprachwissenschaft, 11–101. München: Fink. [Originally published as Sistema,
norma y habla, Montevideo 1952.]

Himmelmann, Nikolaus P., 1996, Zum Aufbau von Sprachbeschreibungen. Linguistische
Berichte 164:315–333.


Kay, Paul and Charles J. Fillmore, 1999, Grammatical constructions and linguistic

and Auckland University Press.

—— 1990, Aps basd skop kmm ak pak ñbelgpal. In Andrew K. Pawley, ed. Kalam hunting

Majnep, Ian S. and Andrew K. Pawley, 2001, On the value of ecological knowledge to
the Kalam of New Guinea: an insider’s view. In Luisa Maffi, ed. On biocultural
diversity: linking language, knowledge and the environment, 343–357. Washington,
DC: Smithsonian Institute.

Pawley, Andrew K., 1993, A language which defies description by ordinary means.
In William Foley, ed. The role of theory in language description, 87–129. Berlin:
Mouton de Gruyter.
Epilogue: taking up the ‘Pawley challenge’


