

# *Issues in Austronesian morphology*

*A focusschrift for Byron W. Bender*

## *Also in Pacific Linguistics*

---

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

John Lynch and Terry Crowley, 2001, *Languages of Vanuatu: a new survey and bibliography*.

Jane Simpson, David Nash, Mary Laughren, Peter Austin and Barry Alpher, eds, 2001, *Forty years on: Ken Hale and Australian languages*.

Catharina Williams-van Klinken, John Hajek, Rachel Nordlinger, 2001, *A grammar of Tetun Dili*.

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

---

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, the Philippines, Indonesia, 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 Editorial Board of Pacific Linguistics is made up of the academic staff of the School's Department of Linguistics. 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.

The 470 publications before Publication 501 were numbered in four series, A, B, C and D, but the series were merged from the beginning of the year 2000 because the rationale for them had gradually disappeared.

FOUNDING EDITOR: Stephen A. Wurm

EDITORIAL BOARD: Malcolm D. Ross and Darrell T. Tryon (Managing Editors), I Wayan Arka, John Bowden, Andrew Pawley, Paul Sidwell.

Pacific Linguistics 519







# Issues in Austronesian morphology

A focusschrift for Byron W. Bender

edited by

Joel Bradshaw and Kenneth L. Rehg



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

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

Copyright © The authors

First published 2001

National Library of Australia Cataloguing-in-Publication entry:

Issues in Austronesian morphology: a focusschrift for Byron W. Bender.

Bibliography.  
ISBN 0 85883 485 5.

I. Austronesian languages – Morphology. I. Rehg, Kenneth L.,  
1939-. II. Bradshaw, Joel. III. The Australian National University.  
Research School of Pacific and Asian Studies. IV. Title  
(Pacific Linguistics 519)

499.2

Sea chart from the Marshall Islands. Adapted by Barbara Pope from James Hornell, *Canoes of Polynesia, Fiji, and Micronesia*, volume 1 of *Canoes of Oceania* (Honolulu: Bishop Museum, 1936) for *The Contemporary Pacific*, vol. 3, no. 1 (Honolulu: University of Hawai'i Press, 1991). Reprinted by permission.

Cover design by Emily Brissenden  
Printed and bound by Union Offset Printers, Fyshwick, Canberra

# Contents

---

Alfred Capelle	Dr. Bender's work on Marshallese	1
George W. Grace	Byron	3
	A Byron W. Bender bibliography	11
Robert Blust	Historical morphology and the spirit world: The *qali/kali-prefixes in Austronesian languages	15
Joel Bradshaw	The elusive shape of the realis/irrealis distinction in Jabêm	75
Terry Crowley	What happened to Erromangan possessive morphology?	87
Sheldon P. Harrison	The Gilbertese <i>-i</i> intransitives, high-vowel erasure, and related phenomena	105
Frantisek Lichtenberk	On the morphological status of thematic consonants in two Oceanic languages	123
John Lynch	The development of morphologically complex possessive markers in the Southern Vanuatu languages	149
Waruno Mahdi	Personal nominal words in Indonesian: an anomaly in morphological classification	163
Andrew Pawley	Proto Polynesian *-CIA	193
Kenneth L. Rehg	Pohnpeian possessive paradigms: the smart solution, the dumb solution, and the Pohnpeian solution	217
Lawrence A. Reid	On the development of agreement markers in some Northern Philippine languages	235
Malcolm D. Ross	Proto Oceanic *i, *qi, and *-ki	259
Wolfgang Sperlich	Semantic and syntactic functions of reduplication in Niuean	279



# *Dr Bender's work on Marshallese*

---

ALFRED CAPELLE

When I first started working with Dr Bender on the Marshallese–English dictionary project in 1971, I really didn't have any idea what I was getting into. Yes, I knew the work had to do with my being a good and helpful informant on the Marshallese language, being someone, as a native speaker, that Dr Bender could question regarding the meaning of words and the acceptability of various sentences or phrases in the language of the Marshall Islands. Still, it was a totally new kind of endeavor for me. The kind of work I had done immediately prior to this was teaching in the elementary schools on Ebon and, later, Likiep atolls.

While working with Dr Bender, I was also required to take linguistics and ESL courses so I could appreciate what was in store for me as a Marshallese language informant. I suppose I had to know, in general terms, how a language works, so that I could understand the significance of the information about Marshallese that I would be sharing with Dr Bender. In retrospect, not only did the linguistics and ESL classes that I took at the University of Hawai'i benefit me greatly, but I also profited enormously from the professional and personal counsel that Dr Bender freely and gladly gave me during the course of our work.

Dr Bender, as many of you may know, spent a good portion of his life working in the Marshall Islands, helping the Marshallese to become better school teachers and administrators. He and his family lived and worked in Majuro. They also spent time working at the Trust Territory of the Pacific Islands headquarters on Guam and Saipan, working for the Department of Education. Dr Bender greatly contributed to the improvement of the educational system in the islands during the time he and his family lived and worked with us in the 50s and 60s.

When I first met Dr Bender I readily became his protégé. He was an important mentor and guide for me while I was working with him on the Marshallese–English dictionary, as well as while I was a student at the University of Hawai'i. He helped me plan my course work in such a way that I was able to complete my undergraduate work at the University in three years and my graduate studies in two years. No time was wasted working with Dr Bender!

I owe much to Dr Bender for his careful and professional guidance and counsel. I continue to visit with him and discuss topics relative to Marshallese language and culture whenever the opportunity presents itself. Dr Bender and I, along with other Marshall Islanders, continue to collaborate in the preservation and perpetuation of the Marshallese language. It is an endeavour that has kept us going in the past and will continue to do so into the future. Dr

---

Joel Bradshaw and Kenneth L. Rehg, eds, *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*, 1–2. Canberra: Pacific Linguistics 2001

© Alfred Capelle

## 2 *Alfred Capelle*

Bender is a man of remarkable knowledge about language, particularly the Marshallese language and its key role in the lives of our island people. I wish him the best.

# Byron

---

GEORGE W. GRACE

Byron's main linguistic interest has always been Micronesian linguistics, most particularly in the Marshallese language and the comparative linguistics of Micronesian as a group. However, more recently a second, quite separate, research interest has emerged. A decade or so ago he found himself assigned to teach the Department's course on morphology—a course that he has since taught many times. The task of preparing the course gradually led him to give serious thought to unresolved questions in the field of morphology. He reviewed what he considered the main ones in a guest lecture in a linguistics class in 1994, and has published the background paper for this lecture on his website (Bender 1994). Anyway, his involvement in these problems has developed to such an extent that, of late, morphology has become a major area of publication for him.

But this is starting the story at the end. The beginning was on August 14, 1929, when Byron Wilbur Bender was born at Roaring Spring, Pennsylvania. His family were Mennonites who operated a dairy farm. His first schooling was in a little two-room school where he could hear the classes of the older children. The small size of the school permitted the teaching to be somewhat individualised, and Byron completed the first three grades in two years. Both of his parents had an interest in education. His mother had been a teacher, and his father had once planned on a career in teaching. Although things hadn't worked out in such a way as to permit him to enter that profession, he was somewhat able to satisfy this interest by teaching Sunday School and engaging in other similar church-related activities.

When Byron was about 10 years of age, the family moved to Elkhart, Indiana, where his father was employed as treasurer of the Mennonite Board of Mission Charities. They lived in, and his mother served as hostess in, a guest house for visiting members of the same organisation.

Byron spent his last two years of high school in a boarding school, Hesston Academy, in Kansas. His parents hoped this would provide a better environment than the local high school during the unsettled times—this was at the height of World War II.

In the summer following his graduation from high school, he had one of the more memorable adventures of his life. He was employed with some 30 other boys to accompany several hundred horses (a considerable number of which died en route) that were being transported to Poland aboard a World War II Liberty Ship. He particularly enjoyed this exposure to the high seas, and toyed with the idea of joining the merchant marine. However, his father persuaded him to give college a try first.

---

Joel Bradshaw and Kenneth L. Rehg, eds, *Issues in Austronesian morphology: a focusschrift for Byron W. Bender*, 3–10.

Canberra: Pacific Linguistics 2001

© George W. Grace

Thus he entered Goshen College in Goshen, Indiana, a Mennonite-affiliated institution. That was in the fall of 1946. He graduated with an English major three years later, in 1949.

After finishing college he went on to graduate studies at Indiana University. It was at Indiana that he first discovered linguistics. He received his MA (in linguistics) in 1950. He had met his future wife, Lois Marie Graber, while at Goshen, and they were married on August 25, 1950.

Byron went to the Marshall Islands in the summer of 1953. It was then part of the Trust Territory of the Pacific Islands, administered by the U.S. Department of the Interior. He was originally hired under an ambitious plan to prepare dictionaries and grammars of all languages in the Trust Territory. The headquarters and staging area were at Fort Rucker in Honolulu, where Lois obtained a secretarial job. She remained in Honolulu when Byron went on to Majuro in the Marshall Islands, since no housing was yet available in the Marshalls. Shortly after he arrived there, many jobs—including his—were cancelled in a sweeping economy move. However, he'd decided he wanted to remain, and an opportunity presented itself when one of the school teachers who had been hired for the coming term refused to come because of the lack of housing. Byron signed on for that job and Lois soon joined him in Majuro.

His first students were of high school age, although the level of studies was somewhat lower than that of high schools in the United States. Teaching in the Marshalls in this period called for a jack-of-all trades. Byron was involved in a wide variety of activities—teaching, training teachers and running a print shop. He even taught a course in celestial navigation, learning (from a book) as he went.

His first two-year contract was followed by two others. During these terms he took on wider responsibilities in the Trust Territory educational program, finally serving from February to May 1959 as Acting Director of Education for the Trust Territory as a whole.

The first three Bender children (Susan, now married and a practicing attorney in Honolulu; Sarah, who is teaching at the university of Iowa and now has two children; and Cathy, a physician in Kailua) were born during their stay in Micronesia. During this period (1953–59) Byron also was able to collect data for his dissertation.

In 1960–62 he served as Assistant Professor of linguistics and anthropology at Goshen College. There he prepared his dissertation, 'A linguistic analysis of the place-names of the Marshall Islands'. His PhD (with a minor in anthropology) was awarded by Indiana University in 1963.

It was also during the period at Goshen that the remaining two Bender children were born (Judy, now a biochemist at Johns Hopkins; and John, now married, with two children, and employed by the *Honolulu Advertiser*).

The Benders returned to the Trust Territory in 1962—this time to Saipan, where Byron served until 1964 as the English Program Supervisor for the Territory and traveled widely throughout the Trust Territory.

In 1964 they came to Honolulu, where Byron was Associate Professor in the English Language Institute at the University of Hawai'i for one year, before moving in 1965 to the Linguistics Department where he has been ever since.

I have one recollection about our experience in hiring Byron that very neatly foreshadows the role he's played since. Unfortunately, that recollection seems to be false. However, it's too apt for me simply to discard it.

What I remember is that when we were in the process of hiring Byron into the Linguistics Department, we received a letter of recommendation from Fred Householder in which Fred said that he would recommend Byron for any department role **including chairman**. I



remember especially that we (Howard McKaughan and I in particular) found this surprising in view of his limited experience as a university faculty member and the fact that there had been no mention of anything like the department chairmanship for him at that time.

Unfortunately, I seem to be remembering something wrong since searches of the University's files have failed to turn up any such letter, and Howard doesn't remember it. I say 'unfortunately' because whatever it was that I'm remembering (something Fred said orally?) was a foreshadowing of much of Byron's later career, as he did in fact become department chairman in 1969 and served nine successive terms before he finally decided to give up the post in 1995. His fairness and his administrative abilities were such that we faculty members found it easiest just to get on with our work and let Byron handle things.

The University evaluates its faculty members in terms of three kinds of activities: teaching, research and service. Although Byron certainly never neglected his teaching or research, what strikes one as particularly exceptional is the extent of his activities that fit best under the 'service' rubric.

His position as department chair led him to assume wider responsibilities in a number of directions. One was the Linguistic Society of America (LSA). The LSA's 1977 Summer Linguistic Institute was hosted by the University of Hawai'i and the East-West Center. Byron played the leading role in the planning and served as Director of the Institute. During this time he served on the LSA Committee on Institutes and Fellowships. Some years later he served as member, and eventually chair, of the LSA Program Committee (and ex officio member of the Executive Committee). Again, in 1995–97, he served the LSA as Parliamentarian.

Byron has also played a central role in leadership of the faculty union of the University—the University of Hawai'i Professional Assembly (UHPA)—throughout most of its history. UHPA was formed in 1974 as a joint affiliate of the American Association of University Professors (AAUP) and the National Education Association (NEA). Byron joined its Board of Directors in 1978, and played a major role in revising its constitution to bring about a harmonious union of these two rather different constituent organisations.

He served four three-year terms on the Board of Directors in two sequences of two (two consecutive terms being all that the Constitution permits). He also was President for six one-year terms—again all that the Constitution permits. (He was the first member to be awarded the title 'President Emeritus'.) He has also served UHPA as Vice-President, Secretary, Chair of the Finance Committee, and as a member of the Executive Committee, the Collective Bargaining Committee, the Negotiating Team, the Nominations and Elections Committee, the Legislative Committee and the Personnel Committee—most for several terms.

Byron has also played a role in both of the national organisations with which UHPA was affiliated. During his service with UHPA, he was elected a member of the AAUP Council and was a delegate to their annual meetings. He has also been on the Standing Committee on Higher Education of the NEA and has contributed to the NEA journal, *Thought and action*.

As president of UHPA from 1983 to 1988, he had a weekly ninety-second spot in a series on KHVH radio called 'Viewpoint'. In this, he gave his thoughtful views on a wide variety of matters related to the University and its place in the community. A number (103, to be exact) of these were published by UHPA in 1985 as a booklet entitled 'University Views'.

J.N. Musto, Executive Director of UHPA, recommended to the Governor of the state, John Waihee, that Byron be appointed as a trustee of the Hawai'i Public Employees Health Fund. He served successive four-year terms, and held both the position of Vice-Chair of the Trustees and that of Chair of the Benefits Committee. In addition to all of the preceding

responsibilities, he also served for fifteen years on the University of Hawai'i Faculty Senate, to which he was first elected in 1980. During nine of these years, furthermore, he was a member of the Senate Executive Committee.

When I set out to prepare this piece, I was very conscious of the fact that there were whole domains of Byron's professional service of which I had no firsthand knowledge, and so I tried to find and elicit information from people who had worked with him in various organisations. Of the three national organisations—AAUP, NEA and LSA—I was successful only with the LSA, but I was able to contact a number of people who had worked with him in the Faculty Senate and UHPA. All of those that I talked with were eager to express their appreciation for his work and for him as a person. Their comments clearly evoked the same person whom we in linguistics had become so familiar with, but they did provide some details of his other activities that we didn't have. A number of these comments deserve quoting here.

Many of us were aware that Byron had played a very significant role in the early stages of the development of UHPA, but knew very little more. Some of the comments make it clear that the union was seriously divided between two camps, and that Byron deserves the principal credit for bringing them together into an effective unit; for example:

Byron was the first leader of UHPA to truly unify the Board members from UHM and the CCs. When I started on the Board, the 'AAUP types' sat on one side of the room and the teachers and education community, which resulted from CUPA, sat on the other side. Byron led the changes in our Constitution that created an independent organization called UHPA out of bifurcated organization defined by the merger of 2 national affiliates. It was through Byron that UHPA became an all encompassing union of UH faculty members. (J. N. Musto)

Our AAUP Chapter was active in the Collective Bargaining contest, initially leading on Manoa and later collaborating with others in forming UHPA. I think that it is a significant testimony to Byron's character and leadership that once he was elected president of UHPA, the rivalries between different groups faded away. (Vincent Peterson)

The same ability to calm troubled waters is reported from other contexts; for example, from the Faculty Senate:

I came to admire his moderate approach, concern for fair play, and refined sense of humor which eased tensions when discussions became a bit heated. (Vincent Peterson)

The impression on seeing him in the SEC, or in the Senate, was of an almost extraordinarily quiet man, thinking but perhaps of things other than the topic under discussion. But that was only until he began to speak, when with words so well chosen and thoughtful he brought closure to even the most vociferous of arguments. He had seen both sides of the problem, and he offered a solution, a question, another direction. (Alison Kay)

And from the Linguistic Society of America:

Byron has played many roles in the life of the Linguistic Society—Director of one of the largest and most successful Linguistic Institutes in 1977 (over 800 participants), member of the Society's Program Committee and most recently, the first officially appointed Parliamentarian (1995–1997).

His talents as an able administrator and thoughtful facilitator were a great asset on the Program Committee and in managing the complexities of the Institute. But it was in his role as Parliamentarian that these gifts took on a more public face and his serenity and personal kindness shone. During the 1997 Annual Meeting in Chicago, the national debate on ebonics brought enormous press interest in the LSA Business Meeting where a

resolution on ebonics was presented for the members' endorsement. While television crews jockeyed to capture the essence of a lively, scholarly debate about the resolution, Byron maintained order amid a flood of motions and substitute wordings, bringing unanimity on the statement AND adjourning the session on time. (Maggie Reynolds)

And finally, comments on Byron as a person:

Dr. Bender is a wonderful human being. He is patient and kind, has a sense of humor, is a connoisseur of good food and wine, enjoys fine music, is a good listener ... he is humble and humane. I have never seen him undermine anyone to make them feel inferior. He treats people with dignity and has never expressed any arrogance. (Pamela Tsuru)

(I'll also include here the following unsolicited comment from my wife:)

A thing I notice about Byron—always—is his chuckle—his deep laugh that goes knees to eyes: his wit and warmth are always on his face. (Liz Foster)

All of these comments ring true to those of us who've known him in other contexts.

It's true that he's soft-spoken, but he is capable of making himself heard when he's ready to speak. Certainly, he doesn't at all fit the stereotype of the glib, fast-talking academic. He could be described as a man of few words—one who thinks before he speaks.

In fact, this thinking before speaking is one trait that many of us have found disconcerting. For instance, one would ask him a question, and he'd sit and stare into space for long enough that one would feel increasingly uncomfortable, feeling the need for something to be said, and finally—unable to hold out any longer—repeat or reword the question. Then, almost simultaneously, he'd give his answer—which he'd been thinking out all along.

Not only is he a connoisseur of good food and wine, but he's also an excellent cook. My personal experience of his cooking comes particularly from a period of several years when he and Lois were part of a 'Great Life Dinners' group, the other members of which were Gordon and Anna Fairbanks, Leatrice Mirikitani and me. We met periodically to enjoy one of the dinners from a cookbook entitled, as I recall, *Great Life Dinners*. Each of the dinners consisted of several dishes which, of course, were prepared by the members of the group (with Gordon and me as notable exceptions—our justification was that we 'took care of the drinks'). But it always seemed to me that Byron took responsibility for the most challenging of the dishes, and he always met that responsibility with noteworthy success (and, I thought, considerable enjoyment).

Another interest of Byron's that needs to be mentioned is computers. From the time that personal computers began to assume a significant role in the academic world, Byron effectively took the lead in bringing the department members into the world of the PC and word processing. And ever since he's remained impressively well-informed—more so than almost any of us—on ongoing developments in personal computers and, more recently, the Internet.

Byron has been involved in educational activities throughout his professional life. In addition to his own teaching and his administrative responsibilities for instructional programs at the University of Hawai'i, he has a continuing interest in the people of the Pacific islands and their languages. This interest is represented, for example, by his report on *Linguistic factors in Maori education* (1971) prepared for the New Zealand Council for Educational Research. However, the Marshall Islands and the Marshallese language have always been most central. Although in 1969 he published a book of Marshallese lessons designed for Peace Corps volunteers, the goal of his main efforts has been (to borrow the words of Capelle

and Bender 1996:37), “to make the Marshallese language a viable written medium for use in all tasks of daily living, including its use in the islands’ schools”.

This continuing effort is marked by a number of publications and other activities; see, for example, the correspondence with Alfred Capelle in Capelle and Bender (1996).

But so far I’ve said nothing about what has been most central in his professional life—increasing our knowledge about the languages of the Pacific. He has contributed to this knowledge in a variety of ways in addition to his own research. For one thing he has been editor-in-chief of the journal *Oceanic Linguistics* since 1991 (he had previously served as Managing Editor from 1965, and has been one of the coeditors of the *Oceanic Linguistics Special Publications* since 1965). Under his editorship the journal has blossomed into the kind of vehicle it was intended to be. It has settled into a regular schedule of publication and its contents have more than doubled in size.

I can still quite clearly remember the experience that first made me fully appreciate the quality of Byron’s own linguistic research. It was reading a paper on Marshallese phonology that he had submitted to *Oceanic Linguistics* (where it was published in volume 7, 1968).

The analysis presented in that paper (one which is now quite generally accepted) contained some remarkable conclusions. First of all, he concluded that Marshallese had only a single series—high to low—of vowels. The front–back and rounded–unrounded distinctions that had been reported in earlier studies were reinterpreted as phonologically conditioned variation. But that wasn’t all. The analysis also posited three semiconsonants, *w*, *y*, and *h*, which often were perceptible (at least to English-trained ears) only through their conditioning effects on neighbouring vowels.

Small wonder that when he submitted the paper to *Oceanic Linguistics*, my first reaction (as editor of the journal at that time) was to think that it looked like an exemplary instance of the kind of analysis for which Byron’s mentor, Fred Householder, had coined the name “hocus-pocus linguistics”. Yet, from what I knew of Byron, it was hard to imagine him resorting to hocus-pocus analyses no matter how elegant the results might appear. Now, having known him much longer, I find it even harder to imagine such a thing. Is there **anyone** from whom hocus-pocus would be more out of character?

However, as I read the paper more carefully, I found myself more and more impressed. I was impressed, for one thing, by how clearly Byron wrote. This is something I’ve noted many times since—even in his memos as department chairman. (As chairman he was often frustrated when he sent a draft of some report required by some higher administrative level to the faculty for suggestions and didn’t receive any. I can tell him now that part of the problem was that he expressed things so thoughtfully and clearly that it was hard to think of any improvements to suggest.)

If I found his presentation lucid, I also found it very compelling. It led the reader through the evolution of his analysis, laying out for each successive modification the newly recognised evidence that had determined it. It was not so much an argument for a particular analysis as an explanation of his reasoning, and I could find no basis for questioning any of it.

Byron has numerous other publications on Marshallese. His dissertation was a meticulous and thorough study of place names (the principal results of which were published in Bender 1970). Although this may appear an unusual topic for a dissertation, it’s probably a very good one for involving the investigator at once in many aspects of the language and laying a foundation for further research in a variety of directions. In any event, he has published (sometimes with collaborators) the textbook for an intensive course in the language, a dictionary of “almost 12,000 entries” and several papers on matters of

morphology/morphosyntax. There are also a number of papers that include other Micronesian languages, from survey treatments of Micronesia as a whole to discussions of features shared by Marshallese with one or more others.

And now there is this additional, and quite separate, area of research in which Byron has become increasingly involved. As was mentioned earlier, his having wound up teaching the department's course on morphology led him to give serious thought to unresolved questions in that field.

Of course he had had to deal with problems of morphological analysis in Marshallese from as early as his dissertation, and morphological questions had figured in several of his Marshallese papers, but some of his recent publications have had a significantly different focus. They are concerned with the part played by morphology in human language—with what its bounds are and how it works. The first paper that called this new departure to my attention was Bender 1988 in which he discussed the reduction in the number of conjugations between Latin and Spanish and what kind of analysis might have been able to predict this levelling. Since then, there have been several others—some with no Austronesian content at all. Although there is no question of Byron's having abandoned Marshallese, with the additional time he has gained since leaving his administrative responsibilities he seems to be growing a second parallel career.

#### NOTE

I knew when I undertook to write this piece that I'd have to draw heavily on the knowledge of a number of other people, but I was confident that I'd have no difficulty in getting their assistance. That confidence was completely justified. All of the people whom I approached who had known Byron in one or another capacity proved to be eager to cooperate—to take advantage of this opportunity to express their appreciation for him.

I owe particular thanks to Byron's wife, Lois, who supplied many details of his early life and made a number of helpful suggestions.

In addition, J. N. Musto, Pamela Tsuru and Vincent Peterson all contributed to my understanding of the role Byron played in the history of UHPA—both in the formation and subsequent functioning of the Union itself and in services to the wider community that arose from his UHPA role.

Alison Kay and Vincent Peterson contributed their recollections of his role in the Faculty Senate and in its Executive Committee.

Maggie Reynolds was the only source I was able to find on his activities in national organisations, but she was a good one.

Finally, I should mention that I've also benefited from discussions with Don Topping, Ken Rehg and Alfred Capelle, and that I've received help in identifying and contacting my sources from a number of people who will remain nameless mainly because I didn't think to note down their names.

My sincere appreciation to all of the above, as well as to any others whose contributions I may have overlooked.

#### References

Bender, Byron W., 1968, Marshallese phonology. *Oceanic Linguistics* 7:16–35.

1969, *Spoken Marshallese: an intensive course with grammatical notes and glossary*.  
University of Hawai'i Press.

- 1970, An Oceanic place-name study. In S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic studies in honour of Arthur Capell*, 165–188. Canberra: Pacific Linguistics.
- 1971, *Linguistic factors in Maori education—a report*. Wellington: New Zealand Council for Educational Research.
- 1985, *University views*. Honolulu: University of Hawai‘i Professional Assembly.
- 1988, Predicting morphological change. In Roger Hadlich and J.D. Ellsworth, eds *East meets West: homage to Edgar C. Knowlton, Jr.*, 17–31. Honolulu: Department of European Languages and Literature, University of Hawai‘i.
- 1994, Morphological paradoxes. <http://www2.hawaii.edu/~bender/paradox.html>.
- Capelle, Alfred and Byron W. Bender, 1996, Dealing with the ABCs of Marshallese over twenty years. In France Mugler and John Lynch, eds *Pacific languages in education*, 36–75. Suva: Institute of Pacific Studies, University of the South Pacific.

# *A Byron W. Bender bibliography*

---

(as of December 2000)

- 1952, With Zdenek Salzman and Giuseppe Franciscato. Friulian phonology. *Word* 8(3):216–223.
- 1962, Individual research activities. *Oceanic Linguistics* 1:12.
- 1963, A linguistic analysis of the place-names of the Marshall Islands. PhD dissertation, Indiana University.
- 1963, Individual research activities. *Oceanic Linguistics* 2:80.
- 1963, Marshallese phonemics: labialization or palatalization? *Word* 19(3):335–341.
- 1964, Individual research activities. *Oceanic Linguistics* 3:269.
- 1964, Pretences in language teaching. *English Language Teaching* 19(1):6–13. Reprinted 1965 in *The N[ational] A[ssociation of] T[eachers of] E[nglish] Bulletin* [Republic of China] 1(2):89–96. Reprinted 1968 in *Journal of English Teaching* [Oxford University Press K.K., in association with the British Council in Japan] 2(2):22–32.
- 1965, With Gregory J. Trifonovitch. *A manual for teachers of English in the Trust Territory of the Pacific Islands*. Saipan: [Office of the High Commissioner].
- 1966, Review of Paul Carano and Pedro Sanchez, *A complete history of Guam* (Tuttle, 1964) and of Charles Beardsley, *Guam, past and present* (Tuttle, 1964). *Mennonite Weekly Review*, 24 March, 4.
- 1967, Review of Leonard Newmark, Jerome Mintz and Jan Lawson Hinely, *Using American English* (New York: Harper and Row, 1964). *Language Learning* XVII(ii & iv):207–214.
- 1968, Marshallese phonology. *Oceanic Linguistics* 7(2):16–35.
- 1969, The other orthographic principle: internal reconstruction in Micronesian languages. In Bruce Sommer, Paul Li, Aleli Agbayani, Videa de Guzman, Anne King and Emily Hawkins, eds *Linguistic Society of 1968–1969 [Proceedings]*, 38–48.
- 1969, Vowel dissimilation in Marshallese. *Working Papers in Linguistics* [Department of Linguistics, University of Hawai‘i] 1(1):88–95.
- 1969, *Spoken Marshallese: an intensive course with grammatical notes and glossary*. Honolulu: University of Hawai‘i Press.
- 1970, An Oceanic place-name study. In S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic studies in honour of Arthur Capell*, 165–188. Canberra: Pacific Linguistics.



- 1971, Micronesian languages. In Thomas A. Sebeok, ed. *Current trends in linguistics*, vol. 8: *Linguistics in Oceania*, 426–465. The Hague: Mouton.
- 1971, *Linguistic factors in Maori education*. Wellington: New Zealand Council for Educational Research.
- 1973, Parallelisms in the morphophonemics of several Micronesian languages. *Oceanic Linguistics* 12(1–2):455–477.
- 1973, With Ho-min Sohn. *A Ulithian grammar*. Canberra: Pacific Linguistics.
- 1976, With Takaji Abo, Alfred Capelle and Tony DeBrum. *Marshallese-English dictionary*. PALI Language Texts: Micronesia. Honolulu: University Press of Hawai‘i.
- 1978, Review of Francois-Xavier Nicolas Zewen, *The Marshallese language: a study of its phonology, morphology, and syntax* (Berlin: Dietrich Reimer, 1977). *Journal of the Polynesian Society* 88(1):118–120.
- 1981, A fossilized article in Marshallese. In Jim Hollyman and Andrew Pawley, eds *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 209–228. Auckland: Linguistic Society of New Zealand.
- 1981, Micronesia/la Micronésie. In Heinz Kloss and Grant C. McConnell, eds *Linguistic composition of the nations of the world 4: Oceania/Composition linguistique des nations du monde 4: L’Océanie*, 41–53. Publications of the International Center for Research on Bilingualism, Series E. Québec: Les Presses de l’université Laval.
- 1982, Obituary of Ruth Crymes. *L[inguistic] S[ociety of A[merica] Bulletin* 97:10.
- 1982, Review of Samuel H. Elbert and Mary Kawena Pukui, *Hawaiian Grammar* (Honolulu: University Press of Hawai‘i, 1979). *Journal of the Polynesian Society* 91(4):73–76.
- 1984, Comments on Mac Marshall, Structural patterns of sibling classification in island Oceania: implications for culture history. *Current Anthropology* 25(5):625–626.
- 1984, Editor. *Studies in Micronesian linguistics*. Author. Object marking in Marshallese, 443–465. With Ho-min Sohn, Surface and deep contrasts in Ulithian phonology, 201–212. Canberra: Pacific Linguistics.
- 1984, With Robert W. Hsu, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel and Judith W. Wang. Micronesian cognate sets. Computer printout. Honolulu: Department of Linguistics, University of Hawai‘i.
- 1985, *University views*. Honolulu: University of Hawai‘i Professional Assembly.
- 1985, With Judith W. Wang. The status of Proto-Micronesian. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 53–92. Canberra: Pacific Linguistics.
- 1988, Response: evaluation at the University of Hawai‘i (Symposium on Post-Tenure Evaluation). *Thought & Action* 4(1):61–64.
- 1988, Predicting morphological change. In Roger Hadlich and J.D. Ellsworth, eds *East meets West: homage to Edgar C. Knowlton, Jr.*, 17–31. Honolulu: Department of European Languages and Literature, University of Hawai‘i.



- 1990, With Kenneth L. Rehg. Lexical transfer from Marshallese to Mokilese: a study in intra-Micronesian borrowing. *Oceanic Linguistics* 29:1–26.
- 1991, On the category distributive. In Robert A. Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 11–26. Canberra: Pacific Linguistics.
- 1993, Book notice: Peter Fenn, *A semantic and pragmatic examination of the English Perfect* (Tübingen: Gunter Narr, 1987). *Language* 69:205.
- 1993, Book notice: David Birdsong and Jean-Pierre Montreuil, eds, *Advances in Romance linguistics* (Dordrecht: Foris, 1988). *Language* 69:193–194.
- 1994, Language reform in a Micronesian context. In István Fodor and Claude Hagège, eds *Language reform: history and future*, vol. 6, 86–99. Hamburg: Helmut Buske.
- 1995, Notes from the field: Lend me your ears! *Oceanic Linguistics* 34:226–232.
- 1996, With Alfred Capelle. Dealing with the ABCs of Marshallese over twenty years. In France Mugler and John Lynch, eds *Pacific languages in education*, 36–75. Suva: Institute of Pacific Studies, University of the South Pacific.
- 1996, Distinguishing between inflection and derivation. In Bernd Nothofer, ed. *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, 199–207. Hamburg: Abera Verlag Meyer & Co.
- 1997, Fred Walter Householder. [An obituary]. *Language* 73:560–570.
- 1998, Markedness and iconicity: some questions. In Anna Siewierska and Jae Jung Song, eds *Case, typology, and grammar: in honour of Barry J. Blake*, 57–70. Typological Studies in Language 38. Amsterdam: John Benjamins.
- 1998, The sign gravitates to the word. In Mark Janse, ed. *Productivity and creativity: studies in general and descriptive linguistics in honor of E.M. Uhlenbeck*, 15–26. Trends in Linguistics, Studies and Monographs 116. Berlin: Mouton de Gruyter.
- 2000, Linguistics. In Robert H. Farrell and Joan Hoff, eds *Dictionary of American history: Supplement: AARP–Lyme Disease*, 368. New York: Charles Scribner's Sons.
- 2000, Paradigms as rules. In Videia P. DeGuzman and Byron W. Bender, eds *Grammatical analysis: morphology, Syntax, Semantics*, 14–29. Oceanic Linguistics Special Publication No. 29. Honolulu: University of Hawai'i Press.
- To appear, A perfect strategy for Latin. In Rajendra Singh and Stanley Starosta, eds *Explorations in seamless morphology*. London: Sage Publications.



# *Historical morphology and the spirit world: the \*qali/kali- prefixes in Austronesian languages*

---

ROBERT BLUST

Many Austronesian (An) languages contain evidence of a largely fossilised set of affixes which overlap in form and show no distinguishable difference in meaning/function. Standard procedures of morphological analysis leave us at a complete loss in dealing with this material since (i) the form of the affix is highly variable, although it adheres to a basic pattern which can be generalised as \*qali/kali-, and (ii) the meaning of the affix cannot be inferred by reference to the 'real world' but only through reference to ethnological categories which relate to the fundamental concepts of animism. These data force us to conclude that the well-known phenomenon of doubling, previously attested only with independent morphemes, also occurs under certain conditions with affixes. An analysis of the \*qali/kali- prefixes further highlights the incompleteness of linguistic theories which appeal to universal cognitive principles that are independent of culture.

## **1 Introduction: Chasing butterflies<sup>1</sup>**

One of the most striking features of the lexicon of Proto Austronesian and many of its descendants is the preponderance of disyllabic base forms. Chrétien (1965) found that of 2,216 base forms listed in Dempwolff's (1938) standard comparative dictionary 2,081 (nearly 94%) are disyllabic. Some modifications in reconstructions which have been made since that time lower this figure slightly, but not significantly. Only 98 of the forms counted by Chrétien (about 4.4%) are trisyllabic, and the number of quadrisyllables reaches a grand total of four.

The tendency to disyllabism in Austronesian languages can be illustrated with the numerals and pronouns, since these constitute complete collections of forms which are not arbitrarily

---

<sup>1</sup> An earlier version of this paper with the title 'A linguistic key to the early Austronesian spirit world' was presented as the keynote address at the Third Eastern Conference on Austronesian Linguistics, held at Ohio University in Athens, Ohio, on 6–7 May 1983. For reasons that would be inappropriate to discuss here that manuscript was never published, and it is now my pleasure to offer it in refurbished form to honor Byron W. Bender, who I have known and admired in one capacity or another for over 30 years.

selected from some larger set. If, for the sake of convenience in citing widely distributed forms, we consider the Proto Malayo–Polynesian (PMP) forms rather than the older Proto Austronesian forms that take account of Formosan evidence, we find that the numeral stems 1–10 and 100 are all disyllabic:<sup>2</sup> 1. \*esa/isa, 2. \*duha, 3. \*telu, 4. \*epat, 5. \*lima, 6. \*enem, 7. \*pitu, 8. \*walu, 9. \*siwa, 10. \*sa-ŋa-puluq, 100. \*sa-ŋa-Ratus. A similar pattern appears in the pronouns: \*aku '1SG', \*kahu '2SG', \*ia '3SG', \*kita '1PL INCL', \*kami '1PL EXCL', \*kamu '2PL', \*ida '3PL'. Apart from the occasional trisyllable (\*taliŋa 'ear', \*qabaRa 'shoulder', \*tuqelan/tuqelan 'bone', \*beties 'calf of the leg', \*qapeju 'gall; gall bladder'), body-part terms show the same strong tendency to disyllabism: \*qulu 'head', \*buhek 'head hair', \*bulu 'body hair; feathers', \*daqih 'forehead', \*mata 'eye', \*ijun 'nose', \*ijipen 'tooth', \*dilaq 'tongue', \*qazay 'jaw', \*kulit 'skin', \*liqeR 'neck', \*susu 'breast', \*dahdah 'chest', \*tian 'abdomen', \*pusej 'navel', \*likud 'back', \*qutin 'penis', \*puki 'vulva', \*qaqay 'leg/foot', \*paqa 'thigh', \*qaleb 'knee cap', \*kuhkuh 'fingernail, toenail', \*uRat 'blood vein; tendon', \*daRaQ 'blood', \*pusuq 'heart', \*baRaQ 'lungs', \*qatay 'liver' etc. Again, except for a few exceptional trisyllables (\*bituqen 'star', \*habaRat 'west monsoon'), the familiar pattern appears in terms for the natural environment: \*larŋit 'sky', \*mata ni waRi 'sun' (= 'eye of the day'), \*bulan 'moon', \*quzan 'rain', \*kilat 'lightning', \*harŋin 'wind', \*timuR 'east monsoon', \*taneq/tanaq 'earth', \*batu 'stone', \*daReq 'clay', \*qenay 'sand', \*wahiR 'fresh water', \*tasik 'salt water; sea', \*kahiW 'wood', \*dahun 'leaf', \*buaq 'fruit', \*buŋa 'flower'.

Given this well-established canonical pattern it must come as a surprise when certain semantic categories are often represented by forms which are exceptionally long. The word for 'butterfly' in many Austronesian languages is a case in point. Among Formosan languages both Jeng (1971) and Zeitoun (n.d.) give *talikuan* for the Takbanuad and Takituduh dialects of Bunun, while Ferrell (1969) gives *bulikuan* for an unspecified dialect in the same meaning ('butterfly, moth'). In itself this would be unremarkable, but Paiwan has five terms for 'butterfly', at least three of which appear to be quadrisyllabic or longer: Western Paiwan *kalazuazung* 'large butterfly sp.', *kaliɕungu-ɕunguL*<sup>3</sup> 'generic term for butterfly', *quLipepe* 'butterfly sp.'. The deviation from standard disyllabism continues in Rukai (Tanan dialect) *'aLibaubang*, Rukai (Labuan dialect) *talivavahə*, Amis (Kiwit dialect) *qaLipapang*, Amis (Central dialect) *'adipangpang* and Puyuma (Tamalakaw subdialect of Katipul dialect) *Halivangvang*, Puyuma (Nanwang dialect) *aribanglalawan* 'butterfly, moth'.

Are these unusually long words meaning 'butterfly; moth' peculiar to the Austronesian languages of Taiwan? Even a glance at comparative data for the Philippines shows that they are not. Setting aside transparently reduplicated disyllabic bases such as Kayapa Kallahan *buqbuqlu*, Mamanwa *kabaqkabaq* or Botolan Sambal *pálopálo*, Reid (1971) gives, among others, the following words meaning 'butterfly': Palawan Batak *qalibangbang*, Sarangani

<sup>2</sup> Abbreviations used in this paper include:

Glossing conventions: 1 – first person, 2 – second person, 3 – third person, PL – plural, SG – singular. Subgroup abbreviations: CEMP – Central Eastern Malayo–Polynesian, CMP – Central Malayo–Polynesian, F – Formosan, Oc – Oceanic, SHWNG – South Halmahera–West New Guinea, WMP – Western Malayo–Polynesian; P in front of any of these subgroups indicated Proto, i.e., the protolanguage ancestral to the subgroup. Other notations include: assim. – assimilation, met. – metathesis, sp. – species (singular), spp. – species (plural), unident. – unidentified.

<sup>3</sup> The symbol /L/ is used for a lateral with varying phonetic properties in Formosan languages, in contrast with the voiced lateral liquid //l/. In Dutch sources /dj/, /tj/ and /oe/ have been changed to /j/, /c/ and /u/ respectively, but otherwise the orthography of sources is followed for most languages, except that the glottal stop in Philippine languages generally is written /ʔ/.

Bilaan *kalbangi* (probably from earlier \**kalibangi*), Binukid *kalibángbáng*, Casiguran Dumagat *kalebongbóng*, Gaddang *qalifambang*, Isneg *kulibangbāng*, Itbayaten *kulivaavang*, Cebuano *alibangbáng*, Western Bukidnon Manobo *kelivangbang*, Mansaka *karibangbang*, Sangil *kalibámang*, and Siocon Subanun *kolombangoy*. An inspection of McFarland (1977) and of standard dictionary sources further yields Itawis *alibengbang*, Polillo Dumagat *salibangbang* (Fox 1953:239, fn. 73), Ilokano *kulibangbáng* 'general name for butterfly and moth', Tagalog *alibambáng*, *alibangbáng* 'small roadside yellow-winged butterfly', *aliparó* 'small roadside butterfly, usually with bright yellow wings'.

Continuing our lepidopterous pursuit southward, the following names turn up in Borneo: Timugon Murut *sakuliambang*, Kadazan *tongkulibambang* 'butterfly', *tongkulibambog* 'moth' (with secondary prefixation after fossilisation of \**kuli-*), Kayan *kelebavah* 'moth', Iban *kelebembang* 'moth or butterfly', *kelebumbu* 'butterfly', *kelelawai* 'butterfly sp.', *kele-mambang* 'butterfly', Dusun Malang *kalabamang*, Dohoi *koLobambang*, Murung *tolubambang*. Sulawesi yields another swarm: Sangir *kaliwembang*, Kaidipang *dalibumongo*, Gorontalo *alinua*, Bolaang Mongondow *kalibombang*, *kolibombang*, Banggai *kalitatak*, Bare'e (Pamona) *kalabamba*, Ampana *aliwombo*, Tae' (Southern Toraja) *kalibambang*, *kalubambang*, Uma (West Toraja) *kalibamang*. In Sumatra we find Simalur *alifambang*, and in the Moluccan islands of eastern Indonesia are Soboyo *kalabebang*, Gani *kalibobo* and Buli *aibobang*. Finally, a few stray butterfly names of inferably quadrisyllabic proportions have found their way into Oceania: Gedaged *kilibob* 'butterfly' (also the name of an important culture hero or creator god), Biliau *kalbangbang*, Numbami *kaimbombo*, Bonkovia *kuləmbembe*, Yevali *kulu-mbembe* (Tryon 1976:268), Puluwat *lipwékipwék*, Ponape *lipahrourou*, Trukese *nipwisipwis* 'butterfly'.

What are we to make of this wild chase, and how can we ensure that in the end our observations will amount to more than a mere exercise in butterfly collecting? The first step toward understanding the historical morphology of the foregoing forms and others that will follow, is to organise them into classificatory categories. Clearly, these words share important properties. First, they deviate sharply from the disyllabic canonical shape typical of the vast majority of lexical stems in Austronesian languages. Second, many of them begin with the historical reflex of a sequence \**Cali-*, where \**C* stands here for a consonant of variable shape. This in itself suggests that most or all of these words contain a fossilised affix. But can we say more about them?

Table 1 reorganises the data so as to highlight the reconstructed shapes of the affixal variants generically represented above as \**Cali-* (F = Formosan, WMP = Western Malayo-Polynesian, CMP = Central Malayo-Polynesian, SHWNG = South Halmahera-West New Guinea, Oc = Oceanic). Cognate sets in which both the base and the prefix variant are fully comparable are marked by a preceding numeral, and reconstructions of affixed words are given after the table where these refer to higher-order protolanguages. This serves two purposes: (i) to distinguish independent witnesses from contingent witnesses for the association of \**qali/kali-* with a given semantic category, and (ii) to provide some insight into the antiquity of phonemic variation in the form of the \**qali/kali-* affix. Where a base can be reconstructed in conjunction with an indeterminate form of the \**qali/kali-* affix, this is written \**X-base*.

**Table 1:** Words for 'butterfly' in selected Austronesian languages, isolating historical prefixes

F:	*buli-	Bunun <i>bulikuan</i> (Ferrell 1969:166)
	*kala-	W. Paiwan <i>kalazuazung</i>
	*kali-	Paiwan <i>kalidungudunguL</i>
	*qali-	Rukai (Tanán; Li 1977:3) <i>ʔaLibaubang</i>
	*qaNi-	Amis (Kiwit) <i>qaLipapang</i>
		Amis (Central) <i>'adipangpang</i>
		Puyuma (Tamalakaw) <i>Halivangvang</i>
		Rukai (Ferrell 1969:166) <i>ʔalivavará</i>
	*qari-	Puyuma (Nanwang) <i>aribanglalawan</i>
	*quNi-	Paiwan <i>quLipepe</i>
	*taNi-	Rukai (Labuan) <i>talivavahə</i>
	*tari-	Bunun (Takbanuad, Takituduh) <i>taLikuan</i> (Zeitoun n.d.)
WMP:	*dali-	Kaidipang <i>dalibumongo</i>
	*kala-	1 Bare'e <i>kalabamba</i>
		1 Dohoi <i>koLobambang</i>
		1 Dusun Malang <i>kalabamang</i>
		Subanun (Siocon) <i>kolombangoy</i>
	*kali-	Banggai <i>kalitatak</i>
		Bilaan (Sarangani) <i>kalbangi</i>
		2 Binukid <i>kalibangbang</i>
		3 Bolaang Mongondow <i>kalibombang, kolibombang</i>
		Dumagat (Casiguran) <i>kalibóngbong</i>
		3 Iban <i>kelebembang</i>
		Iban <i>kelebumbu</i>
		Iban <i>kelelawai</i>
		Iban <i>kelemambang</i>
		Kayan <i>kelebavah</i>
		2 Mansaka <i>karibangbang</i>
		2 Sangil <i>kaʔibámbang</i>
		3 Sangir <i>kaʔiwembang</i>
		2 Tae' <i>kalibambang</i>
		Toulour <i>kalipo'po'</i>
		2 Uma <i>kalibamang</i>
		2 Western Bukidnon Manobo <i>kelivangbang</i>
	*kalu-	Tae' <i>kalubambang</i>
		Tae' <i>kalussambang</i>
	*kula-	Bikol <i>kulagbáw</i>
	*kuli-	4 Ilokano <i>kulibangbáng</i>
		4 Isneg <i>kulibangbāng</i>
		4 Itbayaten <i>kulivaavang</i>
		4 Kadazan <i>tong-kulibambang</i>
		Kadazan <i>tong-kulibambog</i>
		4 Timugon Murut <i>sa-kuliambang</i>
	*kuliN-	Malay (Brunei) <i>kulimpapat</i> 'large moth'
	*pali-	Kankanaey <i>palikwáwa</i> (Reid 1971)

	*qali-		Ampana <i>aliwombo</i>
		5	Batak (Palawan) <i>alibangbang</i>
		5	Cebuano <i>alibangbáng</i>
		5	Gaddang <i>alifambang</i>
			Gorontalo <i>alinua</i>
			Itawis <i>alibengbang</i>
			Minangkabau <i>limpapas</i> 'large moth'
		5	Simalur <i>alifambang</i>
		5	Tagalog <i>alibangbáng/alibambáng</i>
			Tagalog <i>aliparó</i>
	*sali-		Dumagat (Polillo) <i>salibangbang</i>
	*talu-		Murung <i>tolubambang</i>
CMP:	*kala-		Misool <i>kalabubun</i> (Wallace 1962:473)
			Sobojo <i>kalabebang</i>
SHWNG:	*kali-	6	Gani <i>kalibobo</i> (Wallace 1962:473)
		6	Buli <i>aibobang</i>
			Mor <i>karimamo'a</i> (Anceaux 1961:37)
Oc:	*kali-	2	Biliau <i>kalbangbang</i>
		6	Gedaged <i>kilibob</i>
		6	Numbami <i>kaimbombo</i>
	*kulu-		Bonkovia <i>kuləmbembe</i>
			Yevali <i>kulu-mbembe</i>
	*qali-		Gilbertese <i>nikanebu</i> 'large moth'
			Ponape <i>lipahrourou</i> <sup>4</sup>
			Puluwat <i>lipwékipwék</i>
			Trukese <i>nipwisipwis</i>

The organisation in Table 1 highlights several facts, including the following:

- (i) The words for 'butterfly' cited here contain a fossilised disyllabic prefix which ranges over at least eighteen partially similar but distinct protoshapes. The number of etymologically independent attestations of each variant appears in parentheses following the form: 1. \*buli- (1), 2. \*dali- (1), 3. \*kala- (5), 4. \*kali- (12), 5. \*kalu- (1), 6. \*kula- (1), 7. \*kuli- (2), 8. \*kuliN- (1), 9. \*kulu- (1), 10. \*pali- (1), 11. \*qali- (9), 12. \*qaNi- (3), 13. \*qari- (1), 14. \*quNi- (1), 15. \*sali- (1), 16. \*tali- (1), 17. \*talu- (1), and 18. \*taNi- (1). Evidence of phonological conditioning which might have given rise to this luxuriantly proliferating allomorphy is completely absent.

<sup>4</sup> The hypothesis that \*qali/kali- words in Nuclear Micronesian languages reflect quadrisyllables that began with \*qali- is speculative. As Ken Rehg has reminded me, apart from \*paka- > \*ka- 'causative' there are few precedents for a claim that initial syllables were ever lost in languages such as Ponapean or Trukese. Given the weak support for monosyllabic prefixes such as \*li-, however, we have little choice but to assume irregular apocope in these forms. Since the most frequent \*qali/kali- variants are \*qali- and \*kali- (Table 5), probability dictates that a choice be made from these allomorphs, and since the frequency of \*qali- and \*kali- as etymologically independent tokens is virtually identical it is arbitrary whether we choose one or the other.

(ii) A number of the forms cited here belong to widely distributed cognate sets. These are marked with numerals 1–6 preceding the forms cited, and are keyed to the following reconstructions (geographically restricted cognate sets are excluded):

1. PWMP \*kala-baŋbaŋ
2. PMP \*kali-baŋbaŋ
3. PWMP \*kali-beŋbaŋ
4. PWMP \*kuli-baŋbaŋ
5. PWMP \*qali-baŋbaŋ
6. PCEMP \*kali-beŋbeŋ

(iii) The variants \*qali- and \*kali- are the most frequent and consequently the most representative members of the set. Iban, Kayan *kele-* (where /e/ is a mid-central vowel) show vocalic neutralisations in prepenultimate syllables which prevent an etymological distinction between variants 3–9 above. Category assignment in these cases and in others which will be discussed below follows probabilities which derive from the frequency of unambiguous reflexes.

(iv) Acehnese *bangbang* ‘butterfly’, reflects the simple base in PWMP \*kala-baŋbaŋ, PMP \*kali-baŋbaŋ, PWMP \*kuli-baŋbaŋ, and PWMP \*qali-baŋbaŋ. Although Acehnese *bangbang* could be secondarily reduced from a longer form, it can be taken at face value as providing contrastive evidence that the longer forms cited here contain a fossilised affix. This conclusion must be reached in any case when we consider the contrasts among the affixes themselves (cf. \*kala-baŋbaŋ, \*kuli-baŋbaŋ, \*qali-baŋbaŋ, where the only common element in the affixes is \*-l-, despite a clear recurrent similarity of form.

Finally, words for ‘butterfly’ that do not contain the \*qali/kali- prefix are often reduplicated, as with Agta *lomlom*, Atta *apo:ppo:q* Inibaloi *boqboqdo*, Tagalog *paruparó*, Mamanwa *kabaqkabaq*, Malay *kupukupu* ‘butterfly’, *ramarama* ‘moth’, Makasarese *kupukupu* ‘kind of large butterfly’, *pallapalla*, *pipipipi* ‘butterfly’, *rararata* ‘kind of large butterfly which one may not kill’, POc \*bebe, Hoava *pepele*, Sengga *pepepele*, Kia *tatala*, ‘Are’are *hepehepe*, Sowa *pulpul*, Lenakel *pwapwauk* ‘butterfly’. There is an undoubted iconicity in such words, which attempt through reduplicative means to capture the restless, haphazard, fluttering motion that so tantalises the eye and captivates the imagination in the flight of a butterfly or moth. Such iconicity probably motivates the fixed reduplication in the base \*baŋbaŋ or its variants \*beŋbeŋ or \*beŋbaŋ. But what motivates the further affixation of \*baŋbaŋ, or many other nonreduplicated bases, with the \*qali/kali- prefix?

## 2 Creepy-crawly creatures, marked and unmarked

As with linguistic phenomena in general, an isolated fact or sets of facts which may seem puzzling when taken alone can sometimes be illuminated through widening the scope of inquiry. Since butterflies and moths are insects, or in folk parlance ‘creepy-crawly creatures’, we might ask whether other creepy-crawly creatures (worms, arachnids, crustaceans and the like) also tend to have exceptionally long names that appear to contain a fossilised prefix of the \*qali/kali- type. A search quickly confirms this hypothesis. Many languages in insular Southeast Asia have distinct terms for two types of leech: (i) the small black leech which clings to the leaves of forest plants and readily transfers to the skin when touched, and (ii) the



much larger dark brown leech with a yellow medial stripe which inhabits rice paddies and attaches to the legs of laborers during the planting, transplanting or weeding of the crop. The forest leech is commonly designated by terms which, like the word for 'butterfly', are exceptionally long, and which begin with a disyllabic onset of \*qali/kali- type. The history of terms for the paddy leech is more complex, but also points to fossilised affixation with \*qali/kali-, as will be seen below.

The canonical deviation of names for the jungle leech is seen in such Formosan terms as Paiwan *Limatjek* 'mountain leech', Rukai (Mantauran) *Limatəkə* 'small ground leech' and Kanakanabu *ʔanimək-a* 'creek leech',<sup>5</sup> in such Western Malayo-Polynesian forms as Isneg (*a*)*limanawan* 'rather large leech, speckled black and yellow', Ilokano *alimátek* 'leech', Tagalog *limátik* 'leech, blood-sucker', Hanunóo *limátuk* 'common leech, bloodsucker, found in damp forests; a bloodsucking annelid worm of the class *Hirudinea*', Cebuano (*a*)*limátuk* 'leech', Maranao *limatek* 'leech', Maranao *salimatek* 'leech', Kadazan *himatok* 'jungle leech', Mukah *selematek* 'leech', Punan Kelai (Antonio Guerreiro n.d.) *lemtak* 'land leech', Singhi *rimotuk* 'land leech', Iban *lemetak* (met.) 'land leech: *Haemadipsa* spp.', Ngaju Dayak *halamantek* 'small forest leech', Banjarese *halimatak* 'kind of insect', Malagasy *dimatika* 'forest leech' (Abinal and Malzac 1963), *dimaty* 'small forest leech' (Richardson 1885), Karo Batak *kalimantek* 'small forest leech', Toba Batak *limatok* 'kind of small leech', Sangir *lamati?* 'leech', Bolaang Mongondow *olimantok* 'leech with blue back and yellow ventral surface', Makasarese *kalimata* 'leech that enters ears or nose', and in such Central Malayo-Polynesian forms as Laora *lamandeka* 'leech', Kambera *lamatak* 'kind of leech: *Haemadipsa* spp.', Roti *kelumatuk* 'leech (generic, but often applied specifically to a very small type)'.

The name of the paddy leech generally is not exceptionally long, since most languages reflect \*qali-metaq, and regular schwa syncope and nasal assimilation in many languages reduced this form to a disyllable apparently reflecting \*lintaq (the reconstruction proposed by Dempwolff). However, a few key witnesses reveal the morphological history of what was clearly a longer word. Names for the paddy leech in Formosan languages include Amis *La-Lintaq* 'mountain leech', and Kanakanabu *niməkəʔə* 'paddy leech', while Western Malayo-Polynesian forms are often reduced to trisyllables or even disyllables: Ilokano *alintá* 'leech', Isneg *alimítá* 'large leech, striped yellow and black', Tagalog *lintá?* 'river leech', Hanunóo *lintá?* 'a carnivorous annelid worm, leech or bloodsucker, found usually in small pools or in bodies of still, fresh water', Cebuano *lintá?* 'leech', Kadazan *himbata* 'water leech' (/h/ from \*l), Punan Kelai *lemta?* 'river leech', Singhi *rimotah* 'water leech', Ngaju Dayak *halaminjau* 'leech used in blood-letting', Malagasy *dinta* 'leech', Iban *lintah* 'buffalo or water leech, *Hirudinaria* spp.', Malay (*ha*)*lintah* 'generic for slow leeches in contrast to springing-leeches (*pachat*)', Jarai *retah* 'water leech', Karo Batak *lintah* 'large water leech', Toba Batak *linta* 'large water leech', Bolaang Mongondow *linta* 'leech', Bare'e *alinta* 'leech', Makasarese *alinta* 'leech'.

Table 2 reorganises the data so as to highlight the reconstructed shapes of the affixal variants as was done in Table 1 for 'butterfly'.

<sup>5</sup> Tsuchida (1976:142) also lists Kanakanabu *ʔaLimək-a* 'paddy leech', but this must be an error, as elsewhere (27ff.) his phoneme inventory for Kanakanabu does not include a voiceless lateral.

**Table 2:** Words for 'leech' in selected Austronesian languages, isolating historical prefixes

I jungle leech (*Haemadipsa* spp.)

F:	*qani-	1	Kanakanabu <i>ʔanimɛək-a</i>
		1	Paiwan <i>Limatjek</i>
		1	Proto Rukai * <i>limatəkə</i>
WMP:	*kali-	2	Karo Batak <i>kalimantek</i>
		2	Makasarese <i>kalimata'</i>
	*qala-	3	Ngaju Dayak <i>halamantek</i>
		3	Sangir <i>lamati'</i>
	*qali-	4	Banjarese <i>halimatak</i>
		4	Bolaang Mongondow <i>olimantok</i>
		4	Cebuano ( <i>a</i> ) <i>limátuk</i>
		4	Hanunóo <i>limátuk</i>
		4	Iban <i>lemetak</i>
		4	Ilokano <i>alimátek</i>
		4	Isneg ( <i>a</i> ) <i>limanáwan</i>
		4	Kadazan <i>himatok</i>
		4	Karo Batak ( <i>a</i> ) <i>limantek</i>
		4	Maranao <i>limatek</i>
		4	Malagasy <i>dimatikaldimaty</i>
		4	Singhi <i>rimotuk</i>
		4	Tagalog <i>limátik</i>
		4	Toba Batak <i>limatok</i>
	*sali-	5	Maranao <i>salimatek</i>
		5	Mukah <i>selematek</i>
	*tali-		Simalur <i>talimata'</i> (Kähler 1961 sub <i>tali</i> 'rope')
CMP:	*kelu-		Roti <i>kelumatuk</i>
	*qala-	3	Kambera <i>lamatak</i>
		3	Laora <i>lamandeka</i>

II paddy leech

F:	*qani-	1	Amis <i>La-Lintaq</i>
		1	Kanakanabu <i>nimɛcaʔə</i>
WMP:	*qala-		Ngaju Dayak <i>halaminjau</i>
	*qali-	2	Bare'e <i>alinta</i>
		2	Bolaang Mongondow <i>linta'</i>
		2	Cebuano <i>lintá?</i>
		2	Hanunóo <i>lintá?</i>
		2	Ilokano <i>alintá</i>
		2	Isneg <i>alimtá</i>
		2	Jarai <i>retah</i>
		2	Karo Batak <i>lintah</i>
		2	Makasarese <i>alinta</i>
		2	Malagasy <i>dinta</i>

	2	Malay ( <i>ha</i> ) <i>lintah</i>
	2	Singhi <i>rimotah</i>
	2	Tagalog <i>lintá?</i>
	2	Toba Batak <i>linta</i>
*qaliN-		Kadazan <i>himbata</i>

The organisation in Table 2 highlights several facts, including the following:

First, the words for 'leech' cited here contain a fossilised disyllabic prefix which ranges over at least seven partially similar but distinct protoshapes. The number of etymologically independent attestations of each variant appears in parentheses following the form: 1. \*kali- (1), 2. \*kelu- (1), 3. \*qala- (2), 4. \*qali- (3), 5. \*qaliN- (1), 6. \*qaNi- (2), 7. \*sali- (1). Since some dictionary citations show variation between a quadrisyllabic base with an initial vowel and a trisyllabic base without this vowel (e.g. Cebuano (*a*)*limátuk*, Karo Batak (*a*)*limantek*; Malay (*ha*)*lintah*), rather than posit additional variants \*Ni-, \*li-, \*la- and the like we will assume that reflexes such as Paiwan *Limatjek* Karo Batak *lintah* or Sangir *lamati?* contain a canonically reduced form of otherwise well-attested disyllabic prefixes \*qaNi-, \*qali-, or \*qala- (cf. the Micronesian words for 'butterfly' in Table 1, where the same applies).

Second, as with Table 1, a number of the forms cited here belong to widely distributed cognate sets. These are marked with numerals 1–5 for the jungle leech, and 1–2 for the paddy leech preceding the forms cited, and are keyed to the following reconstructions. Again, geographically restricted cognate sets are excluded:

1. PAn \*qaNi-matek
2. PWMP \*kali-matek
3. PMP \*qala-matek
4. PWMP \*qali-matek
5. PWMP \*sali-matek
  
1. PAn \*qaNi-meCaq
2. PWMP \*qali-meCaq

Third, a major difference between the words for 'butterfly' and the words for 'leech' is that the former very often contain noncognate bases, whereas most of the latter fall into a small number of cognate sets. One of the results of this contrast is that there is very little difference in the frequency of etymologically independent prefix variants in the words for 'leech'. Thus \*qali- with twenty-eight tokens, and \*qala- with five, are very similar in the number of etymologically independent attestations (three to two). Despite this complication, it is clear from the combination of Tables 1 and 2 that \*qali- and \*kali- continue to be the most frequent variants.

Fourth, the most serious problem in dealing with the material in Table 2 is how to demonstrate that these forms in fact contain a fossilised affix. Dempwolff (1938) reconstructed \*lima(n)tek and \*lintaq, with no hint of morphological complexity in either form. The problem is particularly acute in \*lintaq, since it conforms to the dominant disyllabic canonical shape of most reconstructed morphemes, and many reflexes are irregular if traced instead to \*qali-metaq (e.g. Tagalog *lintá?*). But Dempwolff's reconstruction fails to account for the initial vowel in such widely separated forms as Ilokano *alintá* and Makasarese *alinta*, or the initial syllable in the Malay variant *halintah*. Moreover, it further fails to account for the heterorganic consonant cluster in Isneg *alimta* or Punan Kelai *lemta?*, or to shed any light on the relationship of this cluster to the medial CVC sequence in Singhi *rimotah*. In sum, we are left with no reasonable choice but to posit a base PAn

\*-meCaq, PMP \*-metaq which is most commonly attested in combination with reflexes of an \*qali/kali- affix.

Finally, as with Acehnese *bangbang* 'butterfly', some languages reflect what appears to be an unaffixed form of \*matek 'jungle leech', thus lending further support to the reconstruction of a morphologically complex word \*qali-matek: Bontok *mátek* 'leech', Kankanaey *mátek* 'leech', Ifugaw *mátok* 'leech, bloodsucker', Manggarai *mantek* 'kind of leech: *Haemadipsa* spp.', Ngadha *maté* 'leech'.

Even if we accept the proposal that Dempwolff's \*lima(n)tek, \*lintaq are correctly \*qali-matek, \*qali-meCaq, how can we be sure there is any connection between the fossilised morphology in the words for 'butterfly' and that in the words for 'leech'? The most direct answer to this question is that there is considerable overlap in the shape of the first two syllables of words in Tables 1 and 2. Four of the seven prefix variants isolated in Table 2 (\*qaNi-, \*kali-, \*qali- and \*sali-) also occur in Table 1, and a fifth (\*qaliN-) differs only in the inclusion of a final nasal. In both tables the inferred affix often contains \*N as the second consonant in Formosan languages, and almost invariably contains \*l in this position outside Taiwan. In short, words for both 'butterfly' and 'leech' (two types) in many Austronesian languages not only deviate from canonical norms in pointing to earlier quadrisyllabic shapes, but also contain largely overlapping phonemic material in the first two syllables.

Without further evidence these apparent parallels in the historical morphology of words for 'butterfly' and 'leech' might be dismissed as products of chance. However, a further enlargement of the comparative context leaves no alternative but to conclude that the names of many creepy-crawly creatures contain a fossilised \*qali/kali- affix. In order to make this case convincing it will be necessary to present a substantial quantity of evidence, since otherwise the argument would be vulnerable to charges of selectivity. To save space the remaining examples will be presented in paragraph format. These include terms for ant, bat, beetle, bumblebee, caterpillar, centipede, cockroach, crab, dragonfly, earthworm, firefly, flea, gecko, grasshopper, honeybee, millipede, scorpion, snake, spider, termite, and wasp. Unless otherwise indicated glosses in sources are identical to the headword. Reconstructed shapes of prefixes follow citations in parentheses, and preceding numbers indicate cognate connections, as in Tables 1 and 2. When a cognate base occurs in different highest-order subgroups only with nonidentical \*qali/kali- variants it is reconstructed as \*X + (base), as with PWMP \*X-buyuŋ 'bumblebee'.

#### ANT

F: Siraya *karamoukam* (\*kala-), Saisiat *ʔaLoraʔil* (\*qalu-), Paiwan *quLitsapudus* (\*quNi-) 'large stinging ant', Saisiat *taLopolaeh* (\*talu-)

WMP: Karo Batak *kacirengga* (\*kaci-) 'venomous red biting ant', Maranao *kalalapa* (\*kala-) 'harmless tree ant', Makasarese *kaliwara, kaluara* (\*kali-, with *kaluara* presumably reduced from *kaliwara*), Malay *kelekati, kelekatu* (\*kali-) 'lamp-fly, flying-ant', Pinatubo Negritos (Fox 1953) *kalibóyboy* (\*kali-) 'ant sp.: *Diacama rugosum*', Sangir *kaʔisusu* (\*kali-) 'ant-lion, myrmeleon', Sasak *kaliotong* (\*kali-) 'male flying ant', Toba Batak *halilinga* (\*kali-) 'ant with edible eggs which appears in swarms', Timugon Murut *kalipodos* (\*kali-; cf. \*pejes 'sting, smart') 'fire-ant', Timugon Murut *kaliwata?* (\*kali-) 'small red stinging ants', Karo Batak *kalimpada* 'edible flying white ant' (\*kaliN-), Karo Batak *kalimpagem* (\*kaliN-) 'inedible ant, smaller than *kalimpada*', Malay *kelengkiak* (\*kaliN-) 'bulldog-ant', Angkola-Mandailing Batak *hatinongnong* (\*kati-), Dairi-Pakpak Batak *katikuru* (\*kati-) 'stinging tree ant', Karo Batak *katipiung* (\*kati-) 'reddish-brown ant: *Crematogaster* sp.', Bare'e (1) *lamoti* (\*qala-) 'red tree ant with venomous bite', Bolaang Mongondow

*lojomansik*, (1) *lojomonsik* (\*qala-) 'black ant with very venomous bite', Kankanaey *alalasáng* (\*qala-) 'small red ant', Maranao (1) *lametik* (\*qala-) 'large red ant', Maranao *lamintas* (\*qala-) 'black ant with poisonous sting', Western Bukidnon Manobo (1) *lemetik* (\*qala-) 'generic term for ants', Cebuano *alibusbus* (\*qali-) 'winged large red house ants that come in swarms, esp. during rainy days', Isneg *aliw(alíwāt)* (\*qali-) 'middle-sized, very black, stinging ant', Mori *limonti* (\*qali-) 'ant sp.', Pinatubo Negritos *alilípak* (\*qali-) 'ant sp.: *Solebopsis geminata*', Pinatubo Negritos *alidákdak* (\*qali-) 'ant sp.: *Odontomachus haematoda* Linn.', Proto South Sulawesi \*lintik (\*qali-?; syncope and assimilation) 'ant', Tarakan *linsadam* (\*qaliN-; cf. \*sejem 'ant') 'red fire ant', Kaidipang *lumontiko* (\*qalu-), Kankanaey *atingayáwan* (\*qati-) 'large dark brown ant', Mansaka *atinglá?* (\*qati-; syncope) 'small kind of ant that flies and bites', Maranao *tigasao* (\*qati-) 'tiny red nonbiting ant', Malay *selembada*, *selempada* (\*saliN-) 'large biting ant', Maranao *salimbagat* (\*saliN-) 'flying ant: *Iridomyrmex cordata*', Toba Batak *sarimborbor* (\*sariN-) 'flying white ant that emerges from the ground at night and dies by morning'

CMP: Manggarai *kalawara* (\*kala-) 'small red ant', Soboyo *kalauhong*, *kuhong* (\*kala-), Soboyo *karamoding* (\*kara-) 'large black ant'

Oc: Raluana *kaliloloi*, *kololoi* (\*kali-) 'ant sp.', Trukese *nikúkkútong* (\*qali-) 'ant'

Reconstruction: PWMP \*qala-me(n)tik 'ant sp.' (a base \*metik also is found with other \*qali/kali- variants).

#### BAT

WMP: Bikol *kalabidóng* (\*kala-) 'medium sized bat', Gaddang *kalafíteg* (\*kala-), Malay *kelelawar* (\*kala-; disambiguated by Minangkabau *kalalawa* 'cave bat'), Western Bukidnon Manobo *kelepenit* (\*kala-) 'the smallest local variety of bat', Casiguran Dumagat *kaleputo* (\*kali-) 'winged fruit bat, *Chiroptera* sp.', Gaddang *kalifúteg* (\*kali-), Gaddang *kalifutu* (\*kali-), Tiruray *kelimbungan* (\*kaliN-) 'small fruit bat', Bikol *kulapnúit* (\*kula-) 'small bat', Isneg *kulambág* 'the bat, one of the *Chiroptera*', Balinese *lalawah* (\*qala-; cp. Malay *kelelawar*, Minangkabau *kalalawa*) 'bat sp.; small owl', Karo Batak *alinturu* (\*qaliN-) 'flying fox', Karo Batak (*a*)*lingkaber* (\*qaliN-) 'kind of rather large bat', Toba Batak *ringkabor* (\*qariN-) 'kind of bat', Mukah *selemawa?* (\*sali-) 'flying fox', Kapampangan *talibatab* (\*tali-) 'small bat'

Oc: Ponapean *limwehdi* (\*qali-) 'small sp. of bat'

#### BEETLE

F: Paiwan *quLimamaraw* (\*quNi-) 'iridescent beetle sp.'

WMP: Makasarese *bantimarang* (\*banti-) 'coconut beetle', Karo Batak *kacinangnang* (\*kaci-) 'beetle often found in rotten wood', Tiruray *kelefutey* (\*kala-) 'sweet potato beetle; firefly', Ngaju Dayak *kalambohong* (\*kalaN-) 'black beetle', Angkola-Mandailing Batak *halicungcung* (\*kali-) 'dung beetle', Malay *kumbang kelemata* (\*kali-) 'coconut-beetle' (*kumbang* 'carpenter-bee'), Karo Batak *kalimpenek* (\*kaliN-) 'various scarabaeid beetles or cockchafers', Tiruray *keretarew* (\*kara-) 'taro beetle', Kadazan *hinggaung* (\*qaliN-) 'sago beetle', Hanunóo *alutátip* (\*qalu-) 'a wingless, beetle-like brown insect usually found on the ground', Toba Batak *antingaro* (\*qanti-) 'beetle that destroys the buds of the rice plant', Itawis *asimawá* (\*qati-) 'beetle', Tiruray *tinganga* (\*qati-) 'black coconut beetle: *Oryctes rhinocerus* Linn.', Iban *serentuku* (\*sariN-) 'horned black beetle'

## BUMBLEBEE

WMP: Samihim *kansibuyung* (\*kanti-), Buginese *katimarang* (\*kati-), Itbayaten *alavungan* (\*qala-), Hanunóo *alibúyug* (\*qali-; cp. Tagalog *bu-búyog* 'bumblebee', with partial reduplication but no other affix), Timugon Murut *limumuod* (\*qali-) 'kind of black carpenter bee: *Xylocopa latipes*', Gaddang *alimbafúyug* (\*qaliN-), Gaddang *alimbuyúngen* (\*qaliN-), Ilokano *alimbubúyog* (\*qaliN-), Itawis *arabiyóngen* (\*qara-), Kankanaey (*a*)*timbayúngen* (\*qatiN-), Dusun Deyah *solobuyung* (\*sulu-) 'bumblebee'

Reconstruction: PWMP \*X-buyuŋ 'bumblebee'.

## CATERPILLAR/GRUB

F: Paiwan *quLimamadas* (\*quNi-) 'caterpillar sp.'

WMP: Makasarese *kaluateré* (\*kalu-) 'edible white grub found in coconut, areca and lontar palms', Karo Batak *katimukmuk* (\*kati-) 'long-haired caterpillar that causes itching', Bikol *alaláso?* (\*qala-) 'hairy caterpillar causing an itch where it comes in contact with the skin', Tae' *limara* (\*qali-) 'greenish-yellow caterpillar that causes painful itching', Ilokano *alimbobódo* (\*qaliN-) 'large hairy stinging caterpillar', Ilokano *alimpupúsa* (\*qaliN-) 'thick, soft, white grub, generally living in timber, especially in coco palms, larva of the rhinoceros beetle', Malay *lembata* (\*qaliN-) 'grub of beetle (usually the coconut-beetle) found in decaying palm trunk', Kankanaey *atatádo* (\*qata-) 'bluish-white caterpillar', Itbayaten *antitiris* (\*qanti-) 'caterpillar of a moth'

Oc: Ponapean *limwehdi* (\*qali-) 'caterpillar'

## CENTIPEDE

F: Saisiat (Taai) (1) *ʔaLongæhipan* (\*qalu-), Kavalan *Rusipan* (\*qaru-)

WMP: Makasarese *kalumeme* (\*kalu-) 'reddish creature, half a finger long, with many legs that coils up when touched', Ngaju Dayak *halalipan* (\*qala-?), Banjarese (2) *halilipan* (\*qali-?), Cebuano (1) *aluhípan*, *ulahípan* (\*qalu-; met.), Itbayaten (1) *alipuan* (\*qalu-; met.), (1) Mentawai (1) *alupat* (\*qalu-), Tagalog (1) *aluhípan* (\*qalu-), Tiruray (1) *liyufon* (\*qalu-), Maanyan *anilipan* (\*qani-), Iban *selemada?* (\*sali-) 'forest centipede, black with legs partly white', Hiligaynon *talimbabága* (\*taliN-)

Oc: Mussau (2) *aliéna* (\*qali-?), Motu (2) *aiha* (\*qali-?), Manam (2) *alia* (\*qali-?), Numbami *aluwanga-na* (\*qalu-; Bradshaw 1978:45), Lau (2) *safila* (\*qali-?), Rennellese *'agipaipai* (\*qali-) 'centipede, *Diplopoda* ... considered the embodiment of nonworshipped deities ('apai), and a loathsome creature', Sa'a (1) *áluhe* (\*qalu-)

Reconstructions: 1. PAn \*qalu-Sipan 'centipede'. Note the insertion of a separate phoneme sequence *-ngæ-* between the prefix and the stem in Saisiat, showing that the morpheme division was still recognised at some point in the separate history of this language, 2. possibly PMP \*qali-hipan, but this depends on the interpretation of what appears to be a complex history of metathesis in this form.

## COCKROACH

F: Proto Rukai \*atabaŋə (\*qata-) 'cockroach' (Li 1977:46)

WMP: Bare'e *balabako* (\*bala-), Ifugaw *balaŋ:ngan* (\*bala-; McFarland 1977) 'cockroach', Sangir *baʔakama* (\*bala-), Sangir *baʔukama* (\*balu-), Bare'e *kalapipi* (\*kala-), Bare'e *kalipipi* (\*kali-) 'kind of small cockroach', Makasarese *kulipasa* (\*kuli-) 'kind of large cockroach', Bare'e *alipipi* (\*qali-) 'kind of small cockroach'

## CRAB

F: Paiwan *kaLaviri* (\*kaNa-) 'crab with one large claw and one small' (under *viri* 'left'), Paiwan *qaLačangan* (\*qaNa-) 'large riverine crab', Paiwan *tjibangu* (\*qati-) 'black riverine crab'

WMP: Buginese *kalaumang* (\*kala-) 'hermit crab', Ngaju Dayak *kalapiting* (\*kala-) 'large sea crab' (cf. Malay *kepiting* 'crab'), Sasak *kaliomang* (\*kali-) 'Bernard's crab: *Cenobita bernhardus*', Bolaang Mongondow (1) *olimangow* (\*qali-) 'kind of large crab', Cebuano *alíkúmu?* (\*qali-) 'sea crab with roundish bulging body' (cf. *kúmu?* 'clenched fist'), Cebuano (2) *alimángu* (\*qali-) 'edible crab of tidal swamps', Cebuano *alimásag* (\*qali-) 'edible crab', Chamorro (1) *akmangao*, *atmangao* (\*qali-) 'spotted sea crab', Gaddang (2) *alima:ngu a:ma* (\*qali-) (McFarland 1977:436), Nias (1) *(ali)mango* (\*qali-) 'sea crab', Palauan (1/2) *chemáng* (\*qali-) 'large sea crab', Tagalog (2) *alimángo* (\*qali-) 'large black crab', Ilokano *arimbukéng* (\*qariN-) 'edible crab that burrows in brackish pools along the seashore'

CMP: Manggarai *kalamango* (\*kala-) 'edible sea crab with black body and red claw', Hatusua *lamanu* (\*qala-) 'large crab'

Oc: Arosi (1) *arimango* (\*qali-) 'a very large crab with paddles found in mangrove swamps', Gitua (1?) *alimanga* (\*qali-) 'mud crab', Kove (1) *alimango* (\*qali-) 'mangrove crab', Penchal (2) *kemmíng* 'mangrove crab', Ponapean *likarahs* (\*qali-) 'rock crab', Samoan (1) *alimango* (\*qali-) 'a crab, sp. of *Lupea*' (Pratt 1984), Wuvulu (1) *alimao* (\*qali-) 'crab with large pincer', Ponapean *lisouduhdu* (\*qali-) 'sp. of crab'

Reconstructions: 1. PMP \*qali-maŋaw 'mangrove crab', 2. PMP \*qali-maŋu 'mangrove crab' (doublet of \*qali-maŋaw). Note reflexes of PAn \*qumaŋ 'hermit crab' in both Buginese *kalaumang* and Sasak *kaliomang*.

## DRAGONFLY

F: Saisiat *ʔalʔalyo fayan* (\*qaNi-), Paiwan *quLitsatsengelaw* (\*quNi-)

WMP: Casiguran Dumagat *kalitonton* (\*kali-), Long Terawan Berawan *kariakang* (\*kari-; *akang* = 'ghost'), Kankanaey *alallaóngan* (\*qala-) 'kind of red-brown dragonfly', Bikol *alibangbáng* (\*qali-), Cebuano *alindánaw* (\*qaliN-; *danaw* = 'lake'), Ilokano *alimbubungáw* (\*qaliN-), Cebuano *salindánaw* (\*saliN-)

Oc: Gilbertese *nikanebu* 'dragonfly'

## EARTHWORM

F: Thao *qatidauluk* (\*qati-), Proto Rukai \*atoLiki (\*qatu-), Puyuma (Tamalakaw) *HuRtati* (\*quri-)

WMP: Maranao *kalalanoʔan* (\*kala-) 'earthworm (short, with shiny membranous skin—looks like a small snake)', Binukid *kalimanggid* (\*kali-), Mukah *keleluat* (\*kali-), Mandar *kalindoro* (\*kaliN-), Sangil *lawati* (\*qala-), Sarangani Manobo *eliwati* (\*qali-), Banjarese *halimbatar* (\*qaliN-), Timugon Murut *lingguang* (\*qaliN-), Bikol *aluluntí* (\*qalu-), Bolaang Mongondow (1) *oʔhasi* (\*qalu-), Hanunóo *alukáti?* (\*qalu-), Itbayaten (1) *alwati* (\*qalu-; or \*qali-?), Casiguran Dumagat *alöntayag* (\*qaluN-), Palauan *chulád* (\*quni-) 'earthworm', Pangasinan *alombáyar* (\*qaluN-) 'worm (general term)'

CMP: Roti *kailati* (\*kali-), Ende *taiati* (\*tali-)

Reconstruction: PWMP \*qalu-wati 'earthworm'. For reflexes of the simple base, note Cebuano *wáti* 'earthworm', Uma *wati* 'sago worm'. Roti *kailati*, Ende *tailati* evidently show sporadic loss of prefixal \*l, of the type seen in Buli *aibobang*, Numbami *kaimbombo* 'butterfly'.



## FIREFLY

F: Puyuma (Tamalakaw) *dalipuyupuyan* (\*daNi-; probably with a reflex of \*Sapuy 'fire'), Amis (Kiwit) *qalipunay* (\*qali-), Amis *qalupainai* (\*qalu-), Paiwan *qutsivawvaw* (\*quCi-), Proto Rukai \*taniapoy (\*taNi-; plus \*apoy 'fire')

WMP: Tiruray *kelefutey* (\*kala-) 'sweet potato beetle; firefly', Uma (1) *kalipopo* (\*kali-), Dairi-Pakpak Batak (1) *kalimpetpet* (\*kaliN-), Karo Batak *kalimpétpét* (\*kaliN-), Wolio (1) *kali-kalimpopo* (\*kaliN-; *kalimpopo* = 'star'), Dairi-Pakpak Batak *kalompétpet* (\*kaluN-), Itbayaten *karuaruay* (\*karu-), Makasarese *katiolo* (\*kati-) 'glow-worm', Ilokano *kulalanti* (\*kula-) 'firefly, glow-worm', Ilokano *kulintabá* (\*kuliN-) 'firefly, glow-worm', Malay (Brunei) *kulimpapat* (\*kuliN-) 'moth or firefly', Aklanon *alitáptap* (\*qali-) 'glow-worm—found on rocks exposed at high tide', Bare'e (2) *alipopo* (\*qali-) 'small flying insect, possibly firefly by daylight', Casiguran Dumagat (2) *alipetpet* (\*qali-) 'lightning bug', Isneg (2) *alipatpát* (\*qali-) 'firefly, glow-worm', Tae' *lumpepe* (\*qaluN-), Timugon Murut *andidipot* (\*qandi-), Cebuano *aniniput* (\*qani-), Berawan (Long Terawan) *tebipe?* (\*qati-), Iban (3) *sele(m)pepat* (\*saliN-), Toba Batak (3) *salimpotpot* (\*saliN-)

Reconstructions: 1. PWMP \*kali(m)-petpet 'firefly', 2. PWMP \*qali-petpet 'firefly', 3. PWMP \*salim-petpet 'firefly'.

## FLEA

F: Puyuma (1) *Hatimra* (\*qati-), Saaroa (1) *ʔatimula* (\*qati-), Thao *qatitira* (\*qati-) 'flea (on dog, but not on buffalo)'

WMP: Ibanag *aliffúngo* (\*qali-), Bintulu (1) *temela* (\*qati-), Hanunóo (1) *tímla* (\*qati-) 'common dog flea', Itawis *assímal* (\*qati-), Kankanaey *atílalagá* (\*qati-) 'chicken tick', Western Bukidnon Manobo (1) *tilema* (\*qati-; met.)

Reconstruction: PAN \*qati-mela 'flea'.

## GECKO

F: Bunun *taliNqadaz* (\*taliN-) 'lizard' (possibly not a gecko), Paiwan *qaLalipi* (\*qaNa-) 'gecko lizard', Paiwan *qatjatjipi* (\*qata-) 'gecko lizard (must not be killed, because it is metamorphosed umbilical cord)', Paiwan (Southern dialect) *quNitsatsipi* (\*quNi-) 'gecko lizard'

WMP: Bare'e *kaladidi* (\*kala-), Minangkabau *kalalaso* (\*kala-) 'tree gecko', Isneg *alipāp* (\*qali-) 'the common house lizard. Its cry announces the death of a member of the family', Ilokano *alutiit* (\*qalu-) 'any of the *Lacertilia*; more especially the common house lizard', Kapampangan *lupísak* (\*qalu-)

Oc: Ponapean *limwoahr* (\*qali-) 'gecko'

## GRASSHOPPER

F: Paiwan (Tjavuali dialect) *kaLibungu*, *kaLivungu* (\*kaNi-; Ho 1978:625), Paiwan *tjibungu* (\*qati-) 'grasshopper, cricket' (Ferrell 1982)

WMP: Sundanese *kalicangkas* (\*kali-) 'kind of grasshopper', Sangir *kaḷimbotong* (\*kaliN-), Maranao *karakeban* (\*kara-), Cebuano *alisiwsiw* (\*qali-) 'kind of grasshopper ... not migratory or destructive to crops', Palauan *chebúd* (\*qali-), Maranao *taresik* (\*qata-), Sangir *saḷamangka?* (\*sala-) 'small green grasshopper which makes a chirping noise at night' (= cricket?), Sangir *tuḷakandi* (\*tula-) 'kind of locust'



HONEYBEE (specifically: *Apis indica*)

WMP: Kelabit *berenuan* (\*bari-) 'kind of small bee', Toba Batak *harinuan* (\*kari-) 'kind of large wild bee', Aklanon *ligwan* (\*qali-) 'large horsefly', Bikol *ligwán* (\*qali-) 'bee sp.', Cebuano *ligwán* (\*qali-) 'small wild honeybee having black and orange stripes, nesting inside trees or walls', Hanunóo *alibúbug* (\*qali-) 'black bee with white stripes (family *Anthrophoridae?*)', Tagalog *ligwán* (\*qali-) 'kind of honeybee' (cited sub *anilan*), Kapampangan (Bergaño) (1) *anig-guan* (\*qani-) 'bee sp.', Malay (1) *neruan* (\*qani-) 'a bee or hornet, sp. unident.', Malay (Jakarta) (1) *nyeruan* (\*qani-; assim.) 'bee sp.', Sasak (1) *nyiruan* (\*qani-; assim.) 'kind of bee', Sundanese (1) *nyiruan* (\*qani-; assim.) 'small to middle-sized honey bee that nests in tree hollows and rock clefts', Iban (2) *(re)nyuan* (\*qari-) 'kind of small bee, often kept under eaves in hives of hollowed logs; honey is good, but inferior to that of the large wild *manyi*', Ilokano (2) *arinuán* (\*qari-) 'kind of bee', Isneg (*a*)*ripanggát* (\*qari-) 'small honeyless bee that builds its nest in the stem of light bamboos', Karo Batak *aringgeneng* (\*qariN-) 'a bee: *Apis indica* F.'

Reconstructions: 1. PMP \*qani-Ruan 'a bee, *Apis indica*', 2. PWMP \*qari-ñuan 'a bee, *Apis indica*'. Although a form such as Ilokano *arinuán* might be considered a reflex of \*qani-Ruan with metathesis of the first two consonants, the distinctness of these reconstructions appears to be supported by the occurrence of both bases in unaffixed form: Manggarai *ruang* 'yellowish-red bee, smaller than the ordinary honeybee: *Apis indica*', Ngaju Dayak (2) *bitik nyuan* 'kind of small grey ant', Singhi *nyowan* (\*u > /o/ irreg.) 'house bee'.

## MILLIPEDE

WMP: Tae' *anda* (*kalamoyan*) (\*kala-) 'phosphorescent millipede' (cited in van der Veen 1940 only in the Dutch Register, sub 'duizendpoot'), Balinese *kalimayah* (\*kali-) 'glow-worm', Iban *kelemebai* (\*kali-), Malay (Jakarta) *kalimayah* (\*kali-) 'luminous millipede', Malay *kelemair*, *kelemanyar*, *kelemayar* (\*kali-) 'luminous millipede', Toba Batak *halimontang* (\*kali-) 'phosphorescent light of luminous millipede or toadstool' (van der Tuuk), Malay *kelentugi* (\*kaliN-) 'dark millipede with yellow legs', Karo Batak *katikeran* (\*kati-) 'phosphorescent millipede', Toba Batak *hatitoran* (\*kati-) 'phosphorescent millipede', Banjarese (1) *halimanyar* (\*qali-) 'luminous millipede', Iban (1) *lemayar* (\*qali-) 'small millipede; glow-worm. If crushed at night it gives a light like that of the firefly', Dampelas (1) *alimayar*, *alimemayar* (\*qali-) 'luminous millipede'

Oc: Ponapean *limwakatantar* (\*qali-) 'millipede', Mokilese *limwoskaras* (\*qali-) 'millipede'

Reconstruction: PWMP \*qali-mayaR 'luminous millipede'.

## SCORPION

WMP: Kapampangan *alakdán* (\*qala-), Makasarese *patikala* (\*pati-; cf. Malay *kala* 'scorpion'), Timugon Murut *limpapasa?* (\*qaliN-) 'young scorpion', Mukah *selengatip* (\*sali-) 'scorpion'

## SNAKE

F: Thao *qalimatun* (\*qali-; loan?) 'the umbrella snake or Taiwan banded krait: *Bungarus multicinctus* Blyth'

WMP: Tagalog *alimuranin* (\*qali-) 'large snake sp.', Bikol *alimbusógon* (\*qaliN-) 'green nonpoisonous snake', Ilokano *alindáyag* (\*qaliN-) 'large venomous snake', Isneg *arimarán* (\*qari-) 'black and white venomous snake'

## SPIDER

WMP: Sangir *kaḷabangkang* (\*kala-), Sangir *kaḷibangkang*, *kaḷimangkang* (\*kali-), Sarangani Bilaan *kalmamo?* (\*kali-), Sangir *kaḷubangkang* (\*kalu-), Toulour *karimombot* (\*kari-), Isinay *alingakáwa* (\*qali-), Kankanaey *atingkáwa* (\*qatiN-), Hanunóo *talitágu* (\*tali-) 'poisonous black ground arachnid', Aklanon *talimbabága* (\*taliN-) 'poisonous spider'

## TERMITE

F: Saisiat *ʔaLoraʔil* (\*qalu-)

WMP: Malay *kelekati* (\*kali-) 'lamp-fly, flying ant', Malay *kelekatu* (\*kali-) 'lamp-fly, flying ant', Sasak *kaliotong* (\*kali-) 'male flying ant', Karo Batak Makasarese *alitana* (\*qali-) 'white ant'

## WASP

F: Rukai (Tona) *katiLoLan* (\*kati-; Tsuchida 1976:9)

WMP: Balinese *kalisaswan* (\*kali-) 'kind of wasp.', Dairi-Pakpak Batak *kalipihpih* (\*kali-) 'name given to a certain wasp when it is in its nest; when it leaves its nest to sting someone it is called *endaldal*', Iban *seremukau*, *semukau* (\*sari-)

What are we to make of this bewildering forest of facts? How can we step back and take measure of the whole without becoming lost in the seemingly endless thicket of particulars?

The one thing that appears to be beyond dispute is that many terms for creepy-crawly creatures in Austronesian languages are either quadrisyllabic or reflexes of inferrably quadrisyllabic predecessors. More exactly, such terms frequently contain three, four or more than four syllables. They therefore deviate sharply from the typically disyllabic lexical bases of Austronesian languages. This fact can be illustrated by a syllable count of the semantic categories described above.

**Table 3:** Average number of syllables in words cited for 'creepy-crawly creatures'

WORD	NO. FORMS	NO. SYLLABLES	AVERAGE
1. butterfly	69	280	4.06
2. leech 1	26	95	3.65
3. leech 2	18	48	2.66
4. ant	44	169	3.84
5. bat	16	60	3.75
6. beetle	15	59	3.93
7. bumblebee	11	49	4.46
8. caterpillar	10	42	4.20
9. centipede	19	71	3.74
10. cockroach	8	32	4.00
11. crab	25	92	3.68
12. dragonfly	9	41	4.56
13. earthworm	20	77	3.85
14. firefly	26	106	4.08
15. flea	9	31	3.44
16. gecko	10	38	3.80
17. grasshopper	10	36	3.60
18. honeybee	16	55	3.44
19. millipede	14	57	4.07
20. scorpion	4	15	3.75
21. snake	5	22	4.40
22. spider	9	37	4.11
23. termite	5	20	4.00
24. wasp	4	17	4.25

There are several objections that might be raised against these figures. First, the material cited is **selected** (compare with the ordinary disyllabic generic terms PMP \*sejem 'ant', \*qulej 'maggot; caterpillar', \*kaRaŋ 'freshwater crab', \*wani 'honeybee', \*sisiq/susuq 'edible snail', \*nipay, \*hulaR 'snake', \*lawaq 'spider', or \*anay 'termite', or the trisyllables \*paniki 'fruit bat, flying fox', or \*abuqaŋ 'palm beetle', which appear to have no connection with the \*qali/kali- prefix). Second, the counts do not distinguish cognate from noncognate material, and so inflate the figures for average word length where reflexes of a quadrisyllable are widely distributed. Finally, the counts do not make allowance for reduplication, affixation unrelated to the \*qali/kali- prefixes (for example, the first syllable of Kadazan *tong-kulibambang* 'butterfly', *tong-kulibambog* 'moth', Timugon Murut *sa-kuliambang* 'butterfly'), or the addition of supporting vowels in the historical phonology of some languages, thereby lengthening forms beyond the combination of base + inferred \*qali/kali-affix. Each of these objections can be countered.

With regard to the first objection, many Austronesian languages distinguish a large number of ant species, caterpillar species, crab species, bee species and the like by morphologically unrelated terms. For the Negritos of Mt. Pinatubo, Fox (1953) recorded 20 independent words for ants alone. He was able to obtain Linnaean binominals for fourteen of these, three of which are represented by \*qali/kali- words. The \*qali/kali- terms in such cases apply on the species level, not the genus level. In other words, for reasons we have yet to address,

some species of ants, caterpillars, crabs etc. appear to be morphologically marked, while others are not.

With regard to the second objection the inflation of average syllable length through multiple counting of cognate forms would result in significant differences only if a single quadrisyllabic cognate set was represented in a large number of cases in contrast to many etymologically independent forms which are shorter than four syllables. But none of the distributions fits this pattern. Almost all of the words for 'butterfly', for example, are quadrisyllabic whether or not they are cognate. If we were to collapse all reflexes of \*kala-baŋbaŋ, \*kali-baŋbaŋ etc. into single forms the total number of forms and syllables would be reduced, but the average number of syllables per form would remain virtually unaffected. Essentially the same relationships hold for all of the semantic categories in question, particularly since widespread cognate sets have yet to be established for many of these.

With regard to the last objection it is true that the counts do not make allowance for extrinsic factors which could independently lengthen the forms in question. But this begs the question why such extrinsic factors (at least those which form part of the morphology) would affect these particular semantic categories more than others. With regard to the skewing effects of phonological change the data is drawn mostly from languages which have not undergone canonically altering sound changes. Where such changes have occurred, as in the addition of supporting vowels in Rukai or Malagasy, they are cancelled by languages which have undergone syncope or apocope, as with Palauan, Gedaged or the Nuclear Micronesian languages (for the last see footnote 4). This is particularly clear in reflexes of PAN \*qali-meCaq 'paddy leech', since here both syncope (leading to \*qalimtaq) and apocope (dropping the first syllable) have colluded with nasal assimilation to produce disyllabic forms in many languages that appear innocent of any morphological complication. In conclusion, then, there can be no doubt that the data is representative of real canonical differences in the lexical representation of these semantic categories in comparison with most others.

This is an encouraging beginning, but where do we go next? Any inference beyond this primary conclusion quickly becomes entangled in serious issues of method. Given the dominant disyllabic canonical shape of Austronesian languages the most reasonable conclusion to draw from the quadrisyllabic target of 'creepy-crawly' words is that members of this class contain an affix. There are several pieces of direct evidence that this is the case. First, as noted already, some languages appear to reflect the bare base of a word that is generally attested in \*qali/kali- form: Acehnese *bangbang* 'butterfly', Bontok *mátek*, Manggarai *mantek* 'leech', Cebuano *wáti* 'earthworm', Uma *wati* 'sago worm', Manggarai *ruang* 'yellowish-red bee', Singhi *nyowan* 'house bee'. Similar cases are seen where one language has a disyllabic base and another language a quadrisyllable which incorporates the disyllabic base as its last two syllables, as with Malay *kala*, Makasarese *patikala* 'scorpion'. In some cases, as with Acehnese *bangbang*, the shorter word may be a reduction of the longer form, but in the great majority of cases this is very unlikely given the general phonological development of the languages. Second, in a few cases an \*qali/kali- word contains a well-established disyllabic base which generally occurs without affixation: Timugon Murut *kalipodos* 'fire-ant' (\*pejes 'to sting, smart'), Tarakan *linsadam* 'red fire ant' (\*sejem 'generic for ants'), Buginese *kalaumang* 'hermit crab', Sasak *kaliomang* 'Bernard's crab: *Cenobita bernhardus*' (\*qumaŋ 'hermit crab'). Finally, the very fact that some morphologically complex reconstructions with \*qali-, \*kali- and other variants are supported by comparative evidence and contrast with one another (e.g. \*kali-baŋbaŋ, \*kuli-baŋbaŋ, and \*qali-baŋbaŋ

'butterfly') is evidence that these initial sequences were disyllabic affixes which served to convert disyllabic lexical bases into quadrisyllabic words.

### The problem of hyperallomorphy

Although there is clear support for an \*qali/kali- prefix, recognition of this affix creates two serious problems: (i) we must acknowledge an extraordinary amount of allomorphy (words for 'butterfly' appear to contain at least eighteen inferrably different forms of the prefix which are partially but not completely shared with the words for 'leech'); and (ii) the observed variation does not appear to correlate with phonological or grammatical conditioning. It will be convenient to call this phenomenon 'hyperallomorphy', although the term, which suggests exceptional variability in the shape of a morpheme, is not completely satisfactory. Statistics from a large and globally representative sample of languages are not to hand, but in most well-known languages few morphemes have more than three allomorphs, and scarcely any have more than four. Cases of exceptionally rich allomorphy nonetheless exist, as in Thao of central Taiwan where the actor focus infix *-um/-* has at least eleven surface realisations, although three of these are subphonemic. The crucial difference is that phonological conditions can be stated for the Thao allomorphs of *-um/-*, whereas \*qali/kali-variation shows no obvious conditioning, either phonological or grammatical. On the other hand, a term which stresses the absence of conditioning for \*qali/kali- variants without reference to the richness of variation would fail to distinguish between a morpheme with just two unconditioned allomorphs and the very different case we are observing here.

What hyperallomorphy suggests is a pattern of partially shared history in which the regularity of sound change has been distorted by some factor not normally present in historical development. It is, in short, a kind of affixal equivalent to the problem of doubleting in free morphemes. Gonda (1952) spoke of phonologically and semantically similar free morphemes in many Austronesian languages as constituting 'word families'. With the \*qali/kali- variants we might say we are dealing with an 'affixal word family'. Table 4 provides a frequency count of \*qali/kali- variants in the data summarised in Table 3. In this tabulation cognate sets are counted as single tokens of an affixal variant.<sup>6</sup> Given our usual assumptions (and experience) it is hard to take Table 4 seriously. Has a morpheme ever been reported in any natural language with anything remotely approaching fifty-eight allomorphs? Of course, not all of the variants in Table 4 can be attributed to a single protolanguage. In particular, only a small subset of these variants can be reconstructed for Proto Austronesian. But many others can be attributed to Proto Malay–Polynesian.

<sup>6</sup> In most of the languages cited here reflexes of \*r and \*R are identical. The choice of \*r in affixal variants is determined by a small set of unambiguous reflexes, as with Itawis *arabiyóngen* 'bumblebee', Long Terawan Berawan *kariakang* 'dragonfly', Maranao *karakeban* 'grasshopper', Isneg *arimarán* 'black and white venomous snake' or Toulour *karimombot* 'spider'. It should also be noted that the figures for \*qali- and \*kali- may be somewhat inflated by the assumption that ambiguous reflexes with prepenultimate schwa in Iban, Malay and a few other languages reflect these forms rather than some other less frequent variant.

**Table 4:** Number of etymologically independent tokens of \*qali/kali- prefix variants in the data of Table 3

VARIANT	NUMBER	VARIANT	NUMBER
1. *bala-	4	30. *qali-	46
2. *bari-	1	31. *qaliN-	18
3. *banti-	1	32. *qalu-	11
4. *buli-	1	33. *qaluN-	3
5. *dali-	1	34. *qaNa-	2
6. *daNi-	1	35. *qaNi-	9
7. *kaci-	2	36. *qara-	1
8. *kala-	24	37. *qari-	4
9. *kalaN-	1	38. *qariN-	3
10. *kali-	46	39. *qaru-	1
11. *kaliN-	10	40. *qata-	3
12. *kalu-	4	41. *qanti-	2
13. *kaluN-	1	42. *qati-	15
14. *kaNa-	1	43. *qatiN-	2
15. *kaNi-	1	44. *qatu-	1
16. *kara-	3	45. *quCi-	1
17. *kari-	3	46. *quNi-	6
18. *karu-	1	47. *quri-	1
19. *kanti-	1	48. *sala-	1
20. *kati-	9	49. *sali-	5
21. *kelu-	1	50. *saliN-	4
22. *kula-	3	51. *sari-	1
23. *kuli-	3	52. *sariN-	2
24. *kuliN-	3	53. *sulu-	1
25. *kulu-	1	54. *tali-	5
26. *pali-	1	55. *taliN-	3
27. *pati-	1	56. *talu-	2
28. *qandi-	1	57. *taNi-	2
29. *qala-	14	58. *tula-	2

As a first step toward reducing the complexity of Table 4 we can ignore all variants attested only once. This removes twenty-four items from the list, leaving thirty-four. Second, we can eliminate all variants attested in a single language, regardless of how many times they appear. Finally, in accordance with the well-known problem of unpredictable prenasalisation of medial stops in Austronesian languages we can combine variants such as \*kali- and \*kaliN- or \*qanti- and \*qati- as single forms. In this way we are able to prune the list of fifty-eight variants to perhaps seventeen (1. \*bala-, 2. \*kala-, 3. \*kali-, 4. \*kara-, 5. \*kari-, 6. \*kati-, 7. \*kuli-, 8. \*qala-, 9. \*qali-, 10. \*qalu-, 11. \*qaNi-, 12. \*qari-, 13. \*qata-, 14. \*qati-, 15. \*sali-, 16. \*sari-, 17. \*tali-). At this point further reduction becomes difficult. Ten variants appear in morphologically complex words that have been reconstructed for Proto Austronesian, Proto Malayo-Polynesian or Proto Western Malayo-Polynesian (reconstructions for 'butterfly', 'leech', etc. cited earlier), and all ten of these must be attributed to Proto Western Malayo-Polynesian. Several other variants are widespread, and hence presumably have a long history as affixes, but are not yet reconstructed as components

of morphologically complex words. By far the best attested variants are \*qali- (together with \*qaliN-), with sixty-four etymologically independent examples, and \*kali-, with fifty-six. Other well-attested variants are \*kala- (twenty-five), \*qati- (seventeen), \*qala- (fourteen), \*qalu- (fourteen), \*kati- (ten), \*qaNi- (> PMP \*qani-) (nine), and \*sali- (nine).

Given the partial similarity of most variants it might be argued that the assumed 'prefix' consists of two morphemes which belong to different order classes. The seventeen prefix variants recognised above can be reduced to two order classes with six and eight elements respectively, as follows.

**Table 5:** Hypothetical prefix order classes for the \*qali/kali- affix

ORDER CLASS	
1	2
ba-	la-
ka-	li-
ku-	ra-
qa-	ri-
sa-	ti-
ta-	lu-
	Ni-
	ta-

In certain respects this proposal is the most attractive solution to the problem of hyperallomorphy. Based on the reduced set of seventeen variants recognised above, the order class analysis implies forty-eight combinatorial possibilities, and slightly more than one third of these are realised in the best-attested variants.

Some support for this interpretation may be found in the apparently apocopated reflexes of \*qali/kali- words in some languages, as Hanunóo *limátuk* < \*qali-matek 'jungle leech', Bare'e *lamoti* < \*qala-metik 'ant sp.', or the Nuclear Micronesian languages. Under this interpretation apocope never occurred; rather, the forms in question were affixed with just \*li-, \*la- etc. The problem with this interpretation is that there is very little evidence for the independent prefixation of the **initial** element of an \*qali/kali- prefix (\*\*qa-matek, \*ka-bañbañ etc.). Moreover, even with the considerable reduction of complexity which this analysis permits, apparently free variation remains a formidable problem (reconstructed words for 'butterfly', for example, would contain \*ka-, \*ku- and \*qa-, and reconstructed words for 'jungle leech' would contain \*ka-, \*qa- and \*sa- in the first order class). On balance, then, the case for decomposition of the \*qali/kali- prefix into smaller morphemes does not seem to gain us as much as we need to find a satisfactory solution to the problem of hyperallomorphy.

Alternatively, we might interpret some of the cross-linguistic agreements in \*qali/kali- words as products of convergence, but this also proves difficult to maintain. Straightforward application of the comparative method supports the reconstruction of PWMP \*kuli-bañbañ 'butterfly'. If the apparent cognate set supporting this etymon is actually a product of convergence how much did convergence produce—the entire word, the association of prefix and stem, or just the prefixal variant? It hardly appears plausible that the entire word could arise independently in geographically removed languages. We might then retreat to a hypothesis of convergent association: the evidence for \*kuli-bañbañ arose through independent **association** of \*kuli- and \*bañbañ in several widely separated languages. But if the association of prefix variant and stem was essentially a random process in the separate

histories of individual languages which produced occasional convergence we have no explanation why this particular prefixal variant prevailed and not more common ones such as \*kala- or \*qati-, which never co-occur with \*baɽbaɽ.

Third, we might assume instead that \*kuli- arose from a pre-existing model such as \*kali- through independent modifications of the first vowel. While this is certainly possible it raises other questions (for example, why don't we also find reconstructed variants \*keli-, \*kili-?).

Finally, it must be acknowledged that analytic error may sometimes create the appearance of an \*qali/kali- prefix where none exists. Thus Cebuano *kalibúgan* 'confused' might be seen as reflecting \*kali-bugan. The synchronic morphology, however, points instead to *ka-libug-an*, and in any case a form such as this would not increase the recognised allomorphy of \*qali/kali-. Similarly, reduplication is discounted where it creates the appearance of an \*qali/kali- quadrisyllable, as in Wolio *kalidalida* 'restless', from *kalida* + REDUP, or *ka-* + *lida* + REDUP.

As will become clear in considering a semantically wider class of data, no proposal to reduce the number of \*qali/kali- variants appears to be viable. In fact, as more semantic categories are examined the number of reconstructed variants continues to increase. At the same time we must ask whether the proposed \*qali/kali- prefix is a formally well-defined class. In many words the inferred morpheme consists of STOP + a + l + VOWEL (usually i). But in others the same semantic category is represented by a quadrisyllable which attains its length through reduplication or through the addition of two initial syllables that seem to have no connection to the \*qali/kali- set. Do we consider Bolaang Mongondow *tambilogong*, *tombilogong* 'beetle that bores into sago and coconut palms', Aklanon *kamamangí?* 'tiny crab with one large and one small pincer', Puyuma *kasimaray* 'type of small grasshopper' and Bolaang Mongondow *tontolawa* 'spider' (cf. PMP \*lawaq) as \*qali/kali- words despite their greater divergence from typical variants, or are they simply unrelated to the phenomenon under investigation?

A solution to the problem of hyperallomorphy must await a solution to the other major problem connected with the \*qali/kali- affix—the problem of meaning.

### Semantic markedness

Once we accept \*qali/kali- as a probable prefix we are faced with a second problem. Since morphemes are commonly defined as the minimal units of meaning or grammatical function, and bound morphemes must be associated with some definable semantic or grammatical category, we are obliged to give some idea of what \*qali/kali- might have meant or what function it might have fulfilled.

In the case at hand it is tempting to say that \*qali/kali- was an 'animal prefix'. But a moment's reflection shows that this definition is inadequate. First, apart from 'bat', and two other exceptions to be noted later, none of the categories in question is that of a mammal. Rather, the great majority are arthropods (insects, arachnids, crustaceans), worms (earthworms, leeches) or the larvae of metamorphosing insects (maggots, caterpillars, grubs), with a few reptiles (gecko, various snakes) thrown in. Collectively these can be called 'creepy-crawly' creatures. Was \*qali/kali- then a prefix for this more restricted animal category?

Table 6 lists various creepy-crawly creatures, divided into two categories: those that are linguistically marked with the \*qali/kali- prefix and those that are not.



**Table 6:** Creepy-crawly creatures, marked and unmarked with the \*qali/kali- prefix

UNMARKED	MARKED
1. ant (*sejem)	ant spp.
2. bat, fruit (PMP *paniki)	bat, cave
3. beetle (PMP *abuqarj)	beetle spp.
4. caterpillar/maggot (PMP *qulej)	caterpillar spp.
5. cockroach (*lipes)	centipede
6. crab, freshwater (*kaRarj)	cockroach
7. crab, hermit (*qumarj)	crab sp.
8. crab, coconut (PMP *qayuyu)	bumblebee
9. crab, ghost (PMP *kaRuki)	leech, jungle
10. grub, sago (PMP *qabated)	leech, paddy
11. honeybee (PMP *wani)	honeybee sp.
12. horsefly (*larjaw)	butterfly
13. housefly (*lalej)	dragonfly
14. lizard, monitor (PWMP *bayawak)	gecko
15. lobster/shrimp (*qudarj)	grasshopper
16. louse, head (*kuCu)	wasp
17. louse, body (*CumeS)	earthworm
18. mosquito (PMP *ñamuk)	firefly
19. nit (*liseqeS)	flea
20. paddy bug (*bañaw)	millipede (luminous)
21. snail (PMP *sisiq, *susuq)	scorpion
22. snake (*SulaR, PMP *nipay)	snake spp.
23. spider (PMP *lawaq)	spider spp.
24. termite (*aNay)	termite spp.
25. weevil, rice (PMP *bukbuk)	

As can be seen from Table 6, the hypothesis that \*qali/kali- signalled a general category of ‘creepy-crawly creatures’ also encounters problems, since there is no evidence that the prefix occurred with members of the unmarked category. As much as was practical I have tried to align similar categories in the two columns so as to highlight the differences. Where the categories are similar the unmarked category tends to be generic (ant, beetle, caterpillar, snake, spider, termite), while the marked category singles out individual species. Where the categories are similar but are not separated by a generic/specific distinction, the basis for assignment to the marked or the unmarked class is less obvious. We might speculate that creatures in the unmarked category tend to be more mundane (housefly, horsefly, head louse, body louse), while those in the marked category tend to be more exotic (butterfly, dragonfly, firefly, luminous millipede). But what do terms like ‘mundane’ or ‘exotic’ mean when they refer to creatures that are, in any event, objects of everyday experience for people whose traditional lifestyles kept them out of doors in hunting, gathering and gardening activities?

The classification implicit in Table 6 would appear to be based on a different principle than one of familiarity. Although all of the animal categories cited here can be said to refer to ‘creepy-crawly creatures’, the unmarked members are animals that: (i) tend to invade human space (maggot, cockroach, horsefly, housefly, both types of lice, mosquito, nit), or (ii) are economically important either because they are edible (fruit bat, sago grub, monitor lizard, lobster/shrimp), or because they cause damage to human crops or constructions (paddy bug,

termite, rice weevil). By contrast, the marked creatures in Table 6 for the most part have little or no economic importance. In addition, some of them have properties which might be regarded as eerie: the bioluminescence of fireflies and luminous millipedes, the jerky, vertiginous flight of cave bats and butterflies, the nocturnal chirping of a gecko (one of the few linguistically marked creatures that can be said to invade human space, as it hangs in gravity-defying suspension upside-down from the house rafters). Clearly, there is more to the meaning of the \*qali/kali- prefix than is apparent in the gloss 'creepy-crawly creatures'.

### 3 Rainbow, whirlwind and echo: natural processes and prodigies of nature

In order to simplify the problem, the presentation of data so far has been artificially restricted to one semantic class. But \*qali/kali- words are not limited to the names of creepy-crawly creatures. Somewhat surprisingly, the same type of deviation from typical Austronesian canonical shape appears in words for various natural phenomena. Table 7 contains a selected set of names for 'rainbow', 'whirlwind/whirlpool' and 'echo' in various Austronesian languages.

**Table 7:** Words for 'rainbow', 'whirlwind/whirlpool' and 'echo' in selected Austronesian languages, isolating historical prefixes

#### I Rainbow

F:	*bali-	Proto Rukai *baLilawlaw
	*qali-	Thao <i>qariwazwaz</i>
	*qari-	Kavalan <i>RiwaRwaR</i>
		Puyuma (Tamalakaw) <i>HaRiwanes</i>
	*qaNi-	Bunun <i>qanivalval</i> (Jeng 1971)
	*quNi-	Paiwan <i>quLivangeraw</i>
WMP:	*bala-	Casiguran Dumagat <i>balaghari</i>
	*baliN-	Kankanaey <i>balingkáog</i>
	*bula-	Ilokano <i>bullaláyaw</i>
	*kali-	Balinese <i>kaliacah</i> (Panitia 1978)
		Toba Batak <i>halibutongan</i>
	*kati-	Ifugaw <i>katibongálon</i>
	*qali-	Bolaang Mongondow <i>alibobag, olibobag</i>
	*qati-	Bontok <i>atibongálen</i>

#### II Whirlwind/whirlpool

F:	*buli-	Paiwan <i>vuliLawLaw</i> 'whirlwind'
	*qali-	Puyuma <i>Haripusapus</i> 'tornado'
		Puyuma <i>H-em-arisuwasu</i> 'whirl, swirl'
WMP:	*bali-	Bikol <i>balisúʔsúʔ</i> 'whirlpool'
	*dali-	Bontok <i>dalipospos</i> 'whirlwind'
	*kale-	Maranao <i>kalelenokaʔ</i> 'whirlwind'
	*kali-	1 Karo Batak <i>kalisungsung</i> 'whirlwind'
		Sangir <i>kaʔisusu</i> 'whirlpool, whirlwind'
		Toba Batak <i>haliodong</i> 'to eddy (whirlpools, whirlwinds)'

	1	Toba Batak <i>halisungsung</i> 'whirlwind'
*kaliN-	1	Angkola-Mandailing Batak <i>halincungcung</i> 'whirlwind' Malay <i>kelembubu</i> 'eddying wind; whirlwind'
*qali-	2	Bikol <i>alipúros</i> 'whirlwind, cyclone, tornado' Bikol <i>aliwúswús</i> 'squall, whirlwind' Bolaang Mongondow ( <i>a</i> ) <i>limpurow</i> 'whirlwind' Bontok <i>alipospo</i> 'hair spiral; whirlpool' Cebuano <i>alilúyuk</i> 'whirlpool' Casiguran Dumagat <i>alibúno</i> 'whirlpool' Casiguran Dumaga: <i>alibuteg</i> 'whirlpool' Ifugaw <i>alipuwápu</i> 'whirlwind, cyclone' Ilokano <i>ali(b)nóno</i> 'eddy of water, whirlpool'
	3	Ilokano <i>alipugpúg</i> 'whirlwind, eddy' Ilokano <i>aliponó</i> 'eddy, gyrate, spin, whirl' Isneg <i>alibútag</i> 'small eddy of water' Isneg <i>alikóno</i> 'eddy of water'
	3	Isneg <i>alipugpúg</i> 'whirlwind, eddy of air'
	3	Itbayaten <i>alipugpug</i> 'tornado, whirlwind' Old Javanese ( <i>h</i> ) <i>alisyu</i> 'whirlwind' Old Javanese ( <i>h</i> ) <i>aliwawar</i> 'storm; whirlwind' Tagalog <i>alinugnóg</i> 'gyration'
*qaliN-	2	Cebuano <i>alimpúlus</i> 'small whirlwind' Kapampangan <i>alimpuyut</i> 'whirlpool' (Bergaño) Minangkabau <i>alimbubu</i> 'whirlwind' Tagalog <i>alimpuyó</i> 'whirl or eddy (water, wind)'
*saliN-		Malay <i>selembubu</i> 'whirlwind'
*tali-		Kankanaey <i>talibaw?ék</i> 'to eddy'
*taliN-		Tae' <i>talimpuru</i> 'whirlwind'
Oc:	*kali-	Lakalai <i>kalivuru</i> 'tornado, waterspout' Raluana <i>kalivuvur</i> 'whirlwind, waterspout'
	*qali-	Sa'a <i>áילו'a</i> 'to eddy, of the wind' Sa'a <i>áliupu'e</i> 'to swirl, of pools' Gilbertese <i>nimamano</i> 'whirlpool, eddy' Marshallese <i>likapijwewe</i> 'whirlpool'

## III Echo

F:	*qalu-	Thao <i>qalushinaz</i> 'echo' (loan from Bunun?)
WMP:	*kala-	Tagalog <i>kalatuwát</i> 'warbly echo'
	*kaliN-	Kankanaey <i>kalindakéd</i> 'to echo'
	*qala-	Tagalog <i>alatuwát</i> 'warbly echo'
	*qaleN-	Casiguran Dumagat <i>alempanag</i> 'echo; to echo'
	*qali-	Tagalog <i>alimaymáy</i> 'unintelligible echoing sound' Tagalog <i>alingawngáw</i> 'echoing sound'
	*qaliN-	Karo Batak <i>alinggungi</i> 'echo'
	*qalu-	Tagalog <i>alunigníg</i> 'receding end of an echo'
	*qani-	Bikol <i>aniningál</i> 'echo, reverberation'
Oc:	*qali-	Puluwat <i>likáhenwan</i> 'echo; to echo'

As with previous tables, the organisation in Table 7 highlights several facts:

- (i) The words for 'rainbow', 'whirlwind/whirlpool' and 'echo' cited in Table 7 deviate sharply from the typical disyllabic canonical shape of most morphemes in Austronesian languages. The first two syllables of these words appear to reflect a prefix which ranges over a number of partially similar but distinct protoshapes. In 'rainbow' there are eleven variants (ten with conflation of \*bali/baliN-), in 'whirlwind/whirlpool' there are eleven (eight with conflation), and in 'echo' there are seven (six with conflation). The shape of this element corresponds closely to the range of variation for the first two syllables of the quadrisyllabic forms meaning 'butterfly', 'leech', 'ant' and the like.
- (ii) Some of the forms cited here belong to cognate sets, but none of these are widely distributed.
- (iii) Although higher-level reconstructions are not available for the semantic categories of Table 7 some widely distributed forms share the same *base*, allowing the reconstruction of 1) PAn \*X-waRwaR 'rainbow' (Formosan evidence only), 2) PWMP \*X-cuŋcuŋ 'whirlwind/whirlpool' (Philippine and Sumatran Batak evidence), 3) PMP \*X-pudus (Bikol and Lakalai), 4) PMP \*X-pupuR 'whirlwind/whirlpool' (Philippines and Raluana), and 5) PMP \*X-niŋjal 'echo' (Bikol *aningál*, together with Asilulu *ningal* 'echo', and Hiligaynon *aningál* 'have the delusion of hearing a familiar sound (as one's mother's voice)'). In other cases the affix is revealed by contrast with an unaffixed base in the same language, as with Karo Batak *alinggungi* 'echo', which Neumann (1951) cross-references to *gung* 'large copper gong'.

Little would be gained by citing further statistics on the distribution and frequency of \*qali/kali- allomorphs. Suffice it to say that the variants \*qali- and \*kali- again emerge as the most frequent types, and that there is great overlap with the range of variation established for the inferred prefixes in names of creepy-crawly creatures: seven of the ten conflated variants in 'rainbow', four of the eight in 'whirlwind/whirlpool', and four of the six in 'echo' correspond to the seventeen best-established variants at the end of Table 4, the previously isolated variants \*buli- and \*dali- are strengthened from single to double instantiations, and a new PAn/PMP variant \*bali/baliN- is supported by the sets for 'rainbow' and 'whirlwind/whirlpool'. These observations merely reinforce a point already made, namely that the \*qali/kali- prefix exhibits a surprising and so far unexplained range of seemingly unconditioned variation on a common theme.

The more significant challenge that these new data present is how to bridge the semantic gap between 'creepy-crawly creatures', on the one hand, and various natural phenomena on the other. Although it was shown in Table 6 that \*qali/kali- does not mark creepy-crawly creatures as a class, that material was at least compatible with the interpretation that \*qali/kali- marked some subclass of animals. But now even this interpretation appears untenable, since minimally some natural phenomena must be included with some creepy-crawly creatures in a category defined by common linguistic marking. The full range of relevant natural phenomena noted to date includes the following:

## AUREOLE (lunar/solar halo)

WMP: Bontok *baliwengweng* (\*bali-) 'the circle of light around the moon or sun; halo', Bolaang Mongondow *alitudu* (\*qali-) 'lunar or solar halo', Isneg *alibongbóng* (\*qali-) 'lunar halo', Isneg *ar-aribongbóngan* (\*qari-) 'lunar halo'

Oc: Arosi *arikorokoro* (\*qali-) 'the round halo of the moon'

## DUST

WMP: Ilokano *alipága* (\*qali-) 'flake (of fire); soot, dirt. Any foul or filthy substance adhering to something high', Kapampangan *alipugpug* (\*qali-) 'dust' (Bergaño), Tagalog *alikalabók* (\*qali-) 'dust (rising and falling upon surfaces)', Western Bukidnon Manobo *eliyavuk* (\*qali-) 'dust; of dust, to fly or be stirred up', Ilokano *alinápog* (\*qaliN-) 'dust from putrefied wood, flesh etc.; rising from the mortar when pounding rice etc.', Kankanaey *alimpokápok* (\*qaliN-) 'be thrown up, be dusty, as when the wind throws up sand etc.', Makasarese *alimbu'bu* (\*qaliN-) 'dust', Ilokano *atipurápur* (\*qati-) 'cloud of dust', Ilokano *atipokpók* (\*qati-) 'to blow, to rise (dust); to fly (papers etc.) before the wind'

## SHADOW/REFLECTION

WMP: Ngaju Dayak *kalanjungen, kanjungen* (\*kalaN-) 'shadow, shade', Iban *kelemayang* (\*kali-) 'shadow (esp. moving), dim outline, appearance of spirit or ghost(?), reflection', Kankanaey *alalangáw* (\*qala-) 'shadow; shade', Tae' *lalundun* (\*qala-), Bare'e *limbayo* (\*qaliN-), Ilokano *aniníwan* (\*qani-) 'shade, shadow, image', Isneg *aniníwing* (\*qani-) 'shadow, reflection'

CMP: Soboyo *kalanining* (\*kala-) 'mirror'

## SPARKS

WMP: Karo Batak *turtur, kalinturtur* (\*kaliN-) 'sparks of a fire', Aklanon *alipáeok* (\*qali-) 'live ashes, small cinders that fly in the air from a fire (as from a strong fire on a windy day)', Cebuano *aligatu* (\*qali-) 'fiery particles carried off from a fire by the updraft; give off flaming particles', Ilokano *alipága* (\*qali-) 'flakes of fire', Tagalog *alipáto* (\*qali-) 'flying ember; firebrand', Ilokano *arisangásang* (\*qari-) 'flake (of fire), spark', Isneg *um-arisangásang* 'emit sparks' (probably an Ilokano loanword), Iban *selempepai* (\*saliN-) 'burst, fly off in all directions, as sparks or burning bamboo'

## STORM

WMP: Karo Batak *kalimantung* (\*kali-) 'name of a storm wind', Old Javanese *haliwawar* (\*qali-) 'squall (of wind)',<sup>7</sup> Kankanaey *alimbudádbud* (\*qaliN-) 'to storm, of typhoons'

## SUNSHOWER

WMP: Tae' *balinono* (\*bali-) 'the sun encircled by a rainbow', Isneg *mangar aridádat* (\*qari-) 'alternating rain and sunshine'

<sup>7</sup> Gericke and Roorda (1901) also give this as 'whirlwind'.

### Semantic markedness revisited

It has been shown that the \*qali/kali- prefix marks the names of some creepy-crawly creatures, but not others, in effect providing a linguistic indicator of marked and unmarked semantic categories. Table 8 shows much the same pattern with terms for natural phenomena.

**Table 8:** Natural phenomena, marked and unmarked with the \*qali/kali- prefix

UNMARKED	MARKED
1. sun (PMP *mata ni qalejaw)	aureole
2. dust (PMP *qabuk)	dust
3. noise, sound (PMP *buni)	echo
4. rain (*quzaN)	rainbow
5. shadow/reflection (*qaniNu)	shadow/reflection
6. fire (*Sapuy)	sparks
7. west monsoon (*SabaRat)	storm
8. sunshine (*siNaR)	sunshower
9. wind (PMP *haŋin)	whirlwind/whirlpool

Just as Table 6 shows the contrast between linguistically marked and unmarked categories of creepy-crawly life forms, so Table 8 shows a contrast between what might be called 'ordinary natural phenomena' and 'prodigies of Nature': flies are unmarked, but butterflies are marked, rain is unmarked, but rainbows are marked. Again, there is apparent counter-evidence, but closer examination of the data suggests that this may not be counter-evidence at all. The category 'dust' hardly qualifies as a prodigy of Nature, and in fact appears in both columns. But dust may appear in various forms, most notably lying still or moving, as in a wind. In static contexts it is unmarked, but the glosses for at least some marked forms, as Ilokano *atipokpók*, Ilokano *atipurápur*, Kankanaey *alimpokápok*, Tagalog *alikalábók* and Western Bukidnon Manobo *eliyavuk* suggest that the \*qali/kali- prefix marks (or historically marked) the semantic category 'dust in motion'.

Some of the marked semantic categories in Table 8 are rather weakly represented, and may be invalid. But it is unlikely that all are invalid, and so we are confronted with a recalcitrant question: what do butterflies, leeches, luminous millipedes and the like have in common with rainbows, whirlwinds or echos which distinguishes them from most other meanings that are lexically encoded in Austronesian languages? The answer to this question reminds us that problems in linguistic analysis may sometimes be insoluble unless we take account of correlated features of nonlinguistic culture.

## 4 Semantic contagion

The data surveyed above strongly suggests that it will be futile to attempt to understand the meaning of the \*qali/kali- affix in terms of reference to the 'real world', since nothing in the real world appears to connect butterflies, leeches and luminous millipedes (but not, for example, mosquitoes) with rainbows, whirlwinds and echos (but not rain, wind or fire). In fact, the data examined so far only begin to suggest the magnitude of the problem we must confront in attempting to establish a semantic category delimited by the \*qali/kali- prefix.

One complicating factor not yet noted is that \*qali/kali- sometimes marks categories which are connected by metaphor or by a common abstract property. For example, words for

'butterfly, moth' are extended to names of fish in some languages, and to names of plants in others: (i) Amis *'adipangpang* 'tropical fish (generic term); butterfly', Yami *alibangbang* 'flying fish', Casiguran Dumagat *kalibóngbong* 'butterfly; spadefish: *Scatophagus argus*', Cebuano *alibangbang* 'butterfly; butterfly fish: *Chaetodon* sp.', Sangir *kañiwembang* 'butterfly', *kina? kañiwembang* 'pennant fish', Gedaged *kilibob* 'butterfly; a yellow marine fish about 8 inches long', Numbami *kaimbombo* 'butterfly' > *kembombo* 'butterfly fish', Ponapean *lierpwater* 'butterfly fish'; (ii) Pinatubo Negrito *kalibangbáng* 'a tree: *Bauhinia* spp.' (Fox 1953:239 says this name derives from the similarity of the folded *Bauhinia* leaves to the wings of butterflies), Tagalog *alibambáng*, *kalibambáng*, *kulibambáng* 'small stocky tree, the leaves of which are used for flavoring meat and fish' (only the first of these variants also means 'butterfly' in Tagalog; the others appear in Table 1 in association with different languages), Ilokano *alibangbáng* 'a tree whose sour leaves are used for culinary purposes' (does not mean 'butterfly' in Ilokano, but does in Tagalog and other languages), Karo Batak *kalimbangbang* 'a tall, upright tree: *Alangium begonifolium* Baill.' (does not mean 'butterfly' in Karo Batak, but corresponds closely to the form meaning 'butterfly' in many other languages), Simalur *alifambang* 'butterfly; tree sp.'

These extensions of the primary sense 'butterfly' to various plants and fish (including the butterfly fish) appear to be based on accidental physical similarity rather than on some more fundamental shared property. In much the same way, the extension of \*qali/kali- marking to 'sparks' may reflect a perceived similarity with 'firefly' (metaphor). Such extensions of a morphological category based on perceived similarity can be attributed to a phenomenon that we will call 'semantic contagion'. The likelihood that semantic contagion is real in the case of firefly:sparks is strengthened by cases of **semantic shift**, as the word for 'firefly' has come to mean 'star' in several languages of southern Sulawesi (Wolio *kalipopo*, Muna *kolipopo* etc.). Although semantic contagion and semantic shift are different phenomena (the former involving a transfer of morphological marking to a new semantic category, the latter a transfer of meaning to a preexisting morpheme), both depend upon perceived similarity in meaning.

Other examples appear to be based less on metaphor than on the sharing of a common abstract property. Many languages in the Philippines mark words meaning 'restless' with the \*qali/kali- prefix, perhaps abstracting out the distinctive chaotic quality of the flight of butterflies or bats (this sometimes applies to the swarming of insects or people in a crowd, or the death struggle of fish out of water, of frantically flapping fowls, etc.):

#### RESTLESS

WMP: Sangir *kañabaso?* (\*kala-) 'restless, unable to lie still or sit in one place', Bare'e *kaliowa* (\*kali-) 'struggle, as with pain', Bolaang Mongondow *kalikokab* (\*kali-) 'flap the wings', Kankanaey *kalimúgag* (\*kali-) 'nervous, not able to be still', Sangir *kañidadang* (\*kali-) 'restless, unable to sit still in one's seat', Toba Batak *halioto* (\*kali-) 'to swarm, as ants', Uma *kalipuru* (\*kali-) 'struggle, flap the wings', Sangir *kañindasa* (\*kaliN-) 'restless, unable to control oneself', Sangir *kañintoa* (\*kaliN-) 'restless, unable to lie still or sit in one place', Ilokano *kolipagpág* (\*kuli-) 'to flutter, of birds when they are killed', Timugon Murut *kuliapunapun* (\*kuli-) 'to swarm, of bees and wasps', Ilokano *kuripaspás* (\*kuri-) 'writhe in agony', Casiguran Dumagat *alimuséd* (\*qali-) 'restless, anxious to get up and go', Cebuano *alikási* (\*qali-) 'nervously restless', Cebuano *alipasa* (\*qali-) 'for a child lying in bed to be restless due to discomfort, worry etc.', Cebuano *aliwarús* (\*qali-) 'be restless in expectation of something', Cebuano *aliwása* (\*qali-) 'restless because of discomfort or worry', Ifugaw *alibadbád* (\*qali-) 'make all sorts of twisting, wresting, wringing movements (as wrestlers)', Ifugaw *alikudúkud* (\*qali-) 'jump up and down, as mudfishes which have been caught and put

in a bottle', Ilokano *alíkúteg* (\*qali-) 'restless, turbulent, mischievous', Ilokano *alipúdas* (\*qali-) 'restless, unquiet, uneasy', Ilokano *alíptók* (\*qali-; syncope) 'restless, of a roving disposition', Ilokano *aliwegwég* (\*qali-) 'restless, turbulent, mischievous', Kankanaey *alipattó* (\*qali-) 'struggle, flounder, writhe', Kankanaey *alipugá* (\*qali-) 'restless, wakeful; lie awake, have a sleepless night', Tagalog *alisuwág* (\*qali-) 'disquietude', Cebuano *alindángay* (\*qaliN-) 'uneasy in the body because of a slight fever; have a slight fever, be restless', Cebuano *alindasáy* (\*qaliN-) 'uneasy, restless in place one is lying', Ilokano *alimbádaw* (\*qaliN-) 'to turn in sleeping', Ilokano *alimbásag* (\*qaliN-) 'have insomnia, sleeplessness, abnormal wakefulness', Ilokano *alimbayágan* (\*qaliN-) 'death struggle, said of drowning persons, fish out of the water', Tagalog *alimbayáw* (\*qaliN-) 'uneasy', Tagalog *alumpihít* (\*qaluN-) 'wriggling and twisting (due to discomfort or pain)', Ilokano *aribungbúng* (\*qari-) 'surround in throngs, as when viewing something strange', Ilokano *aripaspás* (\*qari-) 'writhe in agony', Ilokano *arisaksák* (\*qari-) 'to flounder, said of fish', Kankanaey *atipanguá* (\*qati-) 'strive, flounder, writhe', Sangir *saľembiga?* (\*saleN-) 'nervous, skittish, shy', Sangir *saľembuhau* (\*saleN-) 'restless, constantly moving, like the *buhau* fish', Ilokano *saliwagking* (\*sali-) 'unquiet, never still', Sangir *saľihiang* (\*sali-) 'restless, not able to keep still', Sangir *taľigagase?* (\*tali-) 'restless'

CMP: Kamarian *kalapesa* (\*kala-) 'to flounder, of fish'

Oc: Ponapean *lierikik* (\*qali-) 'restless, with reference to a child'

Perhaps conceptually connected are \*qali/kali- words in several widely separated languages meaning 'rustle, move in the wind':<sup>8</sup>

#### RUSTLE

WMP: Dairi-Pakpak Batak *kalimosmos* (\*kali-) 'be carried off quickly by the wind', Kankanaey *kalisíkis* (\*kali-) 'to rustle, as when crumpling up paper', Ilokano *alingyánay* (\*qali-) 'move slightly, of grasses, leaves etc. at the passage of a rat, snake etc.', Kankanaey *alikadóng* (\*qali-) 'to rustle', Ilokano *aringgunáy* (\*qariN-) 'move lightly, of grasses etc. in a breeze', Kankanaey *atikúme* (\*qati-) 'to rustle. A sound, as of the feet of children on the floor when romping'

A similar abstraction may lie behind the use of \*qali/kali- to mark the word for 'arch' (from the shape of the rainbow?) in at least two widely separated Philippine languages: Bontok *aligáweg* (\*qali-), Maranao *kalantimon* (\*kalaN-) 'arch'.

The words for 'whirlwind/whirlpool' (already more abstract than the English equivalents, which are lexically distinct) provide by far the richest illustration of how semantic contagion has enlarged the number of semantic categories originally marked by the \*qali/kali- affix. By way of metaphor we find a connection with 'hair whorl/crown of the head', and by abstraction of distinctive quality we find connections with 'summit', 'dizzy', 'sling (for hurling stones)', 'go in circles, 'confused', 'thick smoke', 'turbid', 'far (hence visually obscure)', 'lost', 'loud noise', 'drunk', 'conical' and perhaps other less well-attested semantic categories:

<sup>8</sup> Dempwolff's (1938) \*harubiru 'commotion' (cf. English 'hullabaloo', from Tagalog *halubilo* 'noisy crowd or multitude'), appears superficially to be an \*qali/kali- word, but points to PMP \*h- rather than \*q-. In view of Malay *haru* 'plaguing, annoying, esp. of an evil spirit plaguing an individual; confusing, throwing into disorder', *orang haru biru* 'a rowdy', this word is best treated as a compound which has been widely borrowed from Malay into other languages of insular Southeast Asia.



## HAIR WHORL

F: Puyuma (1) *Haripuduan* (\*qali-) 'whorl of hair on the head', Thao *qaripazu* (\*qali-) 'hair whorl', Atayal (Mayrinax) *qalipugu* (\*qaNi-) 'hair whorl', Paiwan (Western dialect) *qaLimumudan* (\*qaNi-) 'crown of the head', Proto Atayal (Li 1981:285) \*qalipagu (\*qaNi-) 'hair whorl', Saisiat *kaʔ-alipozaʔan* (\*qaNi-) 'hair whorl', Paiwan *qulipapunu* (\*quli-) 'crown of head; top of mountain, peak', Paiwan *quLipapuduan* (\*quNi-) 'crown of head (where hair whorls)'

WMP: Sangir *kaʔisusu* (\*kali-) 'crown of the head, where the hair whorls', Toba Batak *halisung* (\*kali-) 'hair whorl' (apparently an irregular reduction to distinguish it from *halisungsung* 'whirlwind'), Dairi-Pakpak Batak *kalimbubu* (\*kaliN-) 'crown of the head', Toba Batak *halimbubu* (\*kaliN-) 'crown of the head, place of the fontanel', Aklanon (1) *alipudwan* (\*qali-) 'crown of head (place where hair is found in a whirl)', Aklanon *alipúeos* (\*qali-) 'whirl of hair; whirlpool (as in river)', Bontok *alipospos* (\*qali-) 'crown of one's head; hair spiral; whirlpool', Ilokano *aligusgús* (\*qali-), Ilokano *alipuspús* (\*qali-) 'whorl in the hair', Isneg *alipuspús* (\*qali-) 'whorl (in the hair), anywhere except at the crown of the head', Itbayaten *alisoxed* (\*qali-) 'whirlpool; hair whorl; to whirl; dizzy', Aklanon *alimpupúdwán* (\*qaliN-) 'very centre of top of head (where hair is found in a whirl)', Bikol *alimpupúro* (\*qaliN-) 'the part of the top of the head from which the hair appears to spiral out in different directions', Cebuano *alimpúlu* (\*qaliN-) 'crown, part of the skull in the back where the hair forms a whorl; topmost part of a mountain', Isneg *alintutúbo* (\*qaliN-) 'top, tip, highest point; crown (of the head)', Isneg *lintotóxo* (\*qaliN-; apocope) 'top, peak, summit; crown of the head; whirl at the crown of the head', Hanunóo *aripúdwán* (\*qari-) 'hair whorl', Toulour *ririmpuruan* (\*qari-) 'hair whorl', Toba Batak *salimbubu* (\*saliN-) 'crown of the head, place of the fontanel'

Oc: Marshallese *likapijwewe* 'cowlick, hairwhorl' (cf. 'whirlpool')

Reconstruction: PAN \*qali-pudu-an 'hair whorl'.

The transition from 'whirlwind' or 'whirlpool' to 'hair whorl' to 'crown of the head' to 'summit' is anticipated in some of the glosses already given. However, other \*qali/kali-forms are found meaning 'summit', but with no trace of the connections which apparently led to a reflex of \*qali/kali- being associated with this semantic category:

## SUMMIT

WMP: Aklanon *alipungto(h)* (\*qali-) 'topmost part, peak, pinnacle (of tree, mountain)', Bikol *alituktók* (\*qali-) 'acme, apex, crest, peak, summit, top, vertex, ridge', Ilokano *alimpatók* (\*qaliN-) 'top, peak, summit', Ilokano *alintótok* (\*qaliN-) 'top, peak, summit', Kankanaey *alintayók* 'top, summit (of a tree)', Western Bukidnon Manobo *limbuvungan* (\*qaliN-) 'peak of a house' (cf. PMP \*bubuŋ-an), Bikol *arituktók* (\*qari-) 'acme, apex, crest, peak, summit, top, vertex, ridge' (cf. PMP \*tuktuk 'top, crown, summit'), Ilokano *aringgawís* (\*qariN-) 'top, peak, summit'

## DIZZY

F: Thao *qarimuzmuz* (\*qali-) 'whirling sensation', Bunun *qanipasav* (\*qaNi-) 'faint, dizzy', Paiwan *quLimezaw* (\*quNi-) 'dizziness, "seeing stars"', *quLimezav-en* 'be dizzy'

WMP: Iban *belebigau* (\*bali-) 'giddy, spinning (of the head)', Kankanaey *baliwenwen* (\*bali-) 'turn round'; *ma-baliwenwén* 'dizzy', Sangir *baʔintangó* (\*baliN-) 'groggy, and therefore unsteady on one's feet, as someone who has just disembarked from a boat and is getting his "land legs"; totter; dizziness', Toulour *karimembeng* (\*kari-) 'dizzy, dazed, "seeing stars"', Buginese *alippuang-eng* (\*qali-) 'dizzy', Cebuano *alipúlung* (\*qali-) 'dizzy',

Maranao *lipadeng* (\*qali-) 'swoon, faint; dizziness', Ilokano *alindáw* (\*qali-; from \*qali-medaw?) 'dizzy, giddy, affected with vertigo', Kankanaey *alitenténg* (\*qali-) 'stunned; dumbfounded; astounded; deafened; dizzy; giddy', Pangasinan *alimoréng* (\*qali-) 'dizziness; to faint', Bolaang Mongondow (*a*)*limpurow* (\*qaliN-) 'something, such as a withered leaf, that is swept in a swift upward spiral by the hot air rising from a fire, or by a whirlwind; fit of giddiness; whirlwind', Hiligaynon *atipuyúng* (\*qati-) 'dizzy, feel faint', Ilokano *talimúdaw* (\*tali-) 'feel dizzy, giddy'

CMP: Kambera *kalihingu* (\*kali-) 'dizzy'

Oc: Ponapean *lipwongmas* (\*qali-) 'to faint'

Reconstruction: PAN \*X-medaw 'dizzy'.

From 'dizzy' we are led to 'sling (for propelling stones)', 'go in circles', 'confused', 'thick smoke', 'turbid', 'far', 'lost', 'loud noise, noise that sets the head awhirl' and 'drunk':

#### SLING

WMP: Karo Batak *kalibawang* (\*kali-), *kalimbawang* (\*kaliN-) 'a sling, instrument for propelling projectiles', Kapampangan *alibasbas* (\*qali-) 'sound of a sling whirling' (Bergaño), Ilokano *alimbáyung* (\*qaliN-) 'sling. It generally consists of a strip of the limb of a leaf of the buri palm (*sílag*); both ends of the strip are held in the hand, and it is whirled around until, by loosing one end, the missile, which is mostly a stone, is let fly.'

#### GO AROUND IN CIRCLES

WMP: Ilokano *balikawkáw* (\*bali-) 'walk, talk circuitously, as when trying to intercept somebody, or when using circumlocutions', Tae' *kalumpisa* (\*kaluN-) 'go round in circles', Ilokano *alikuwnéng* (\*qali-) 'fly round about', Kankanaey *alibaybáy* (\*qali-) 'run around, turn about, fly round', Ilokano *arinoknók* (\*qari-) 'to pirouette', Ilokano *salikawkáw* (\*sali-) 'walk, talk circuitously, as when trying to intercept somebody, or when using circumlocutions'

#### CONFUSED (of vision, sound, the mind)

WMP: Sangir *buławuhe?* (\*bula-) 'hazy', *me-buławuhe?* 'seen as through a haze, unclear, hazy, vague (of vision)', Kankanaey *buligawgáw* (\*buli-) 'dull, dim; myope, nearsighted', Kankanaey *bulikawkáw* (\*buli-) 'dark, obscure', Kankanaey *bulingetngét* (\*buli-) 'dark, obscure', Kankanaey *bulisengséng* (\*buli-) 'dark, obscure', Balinese *kalimatmatan* (\*kali-) 'hazy vision, seeing wrongly' (Panitia 1978), Javanese *kalimengan* (\*kali-) 'forget', Kankanaey *kalimattáw* (\*kali-) 'forget; remember only confusedly, not to recognise; see only indistinctly', Kapampangan *kalingwan* (\*kali-) 'forget', Makasarese *kalingongo* (\*kali-) 'unintelligible sound (as of someone who mumbles, or a dying man who speaks only with great difficulty)', Angkola-Mandailing Batak *halimbolos* (\*kaliN-) 'forgetful', Bikol *aliwálas* (\*qali-) 'neglectful', Bontok *allílaw* (\*qali-; syncope?) 'to confuse, as to interject a comment which will make a speaker lose his train of thought', Casiguran Dumagat *alimengmeng* (\*qali-) 'be able to see something for just a split second, and then to have it go out of sight or disappear', Cebuano *alipatpát* (\*qali-) 'for the vision to be blurred', Ilokano *alimadámad* (\*qali-) 'hear confusedly, indistinctly; remember vaguely, indistinctly', Ilokano *aliwangáwang* (\*qali-) 'uncertain, vague, as news', Maranao *lipaḷo* (\*qali-) 'forget, forgetful', Tagalog *aligutgót* 'entanglement of thread or the like', Casiguran Dumagat *alintána* (\*qaliN-) 'not to pay attention to', Isneg *ariyangko?* (\*qari-) 'to sound confusedly', Aklanon *salimúáng* (\*sali-) 'confused, groggy', Cebuano (\*sali-) *salimágaw* 'for the vision to be blurred', Tagalog *saligutgót* (\*sali-) 'intricate, complicated' (cf. *aligutgót*, with visual

confusion), Ifugaw *halinduwá* (\*saliN-) 'doubt' (from *duwá* 'two'), Kankanaey *sulimatmát* (\*suli-) 'forget; remember only confusedly, not to recognise; see only indistinctly', Timugon Murut *taliajow* (\*tali-) 'confused'

Oc: Ponapean *liourehre* (\*qali-) 'mutter indistinctly during sleep, be delirious', *limanokonok* (\*qali-) 'absent-minded, forgetful'

Reconstruction: PWMP \*X-matmat 'see or hear indistinctly or confusedly'.

#### THICK SMOKE/STEAM

WMP: Bikol *alisúhos* (\*qali-) 'smoky', Cebuano *alingásu* (\*qali-) 'smoky, get filled with smoke', Hanunóo *alingháv* (\*qali-) 'vapor rising from the ground', Ilokano *alibongóbong* (\*qali-) 'vapor, steam', Ilokano *alibóob* (\*qali-) 'steam', Ilokano *alibúyong* (\*qali-) 'overcast, clouded over', Ilokano *alingásaw* (\*qali-) 'exhalation in the form of vapor, steam etc.', Ilokano *alisugásug* (\*qali-) 'exhalation, steam rising from the ground after rain', Kankanaey *alinebnéb* (\*qali-) 'full of smoke', Kankanaey *alingángew* (\*qali-) 'to fill (with smoke)', Tagalog *alimu?óm* (\*qali-) 'earth vapor', Hanunóo *alinyábu?* 'fog, mist; drizzle, light rain', Kankanaey *talimbóok* (\*taliN-) 'to smoke much, reek much'

Oc: 'Are'are *aripu?o* (\*qali-) 'be full of smoke', Sa'a *álipono* (\*qali-) 'thick, of smoke', Arosi *aribono* (\*qali-) 'thick darkness; thick smoke, a cloud of smoke'

#### TURBID

WMP: Kankanaey *kalibawbáv* (\*kali-) 'to trouble, make thick, make muddy', Ilokano *aributé?d* (\*qari-) 'filth, impurities in unfiltered water'

CMP: Roti *kelupua* (\*kelu-) 'stir up water, muddy up water'

#### FAR/DISAPPEAR

WMP: Angkola-Mandailing Batak *kalimongmong* (\*kali-) 'run far away with something', Casiguran Dumagat *kalikámed* (\*kali-) 'the far side of the mountains (far away, where the mountains end)', Karo Batak *kalimanman* (\*kali-) 'very far', Karo Batak *kalimatmat* (\*kali-) 'very far', Ilokano *alibtók* (\*qali-; syncope) 'disappear, pass from view'

#### LOST

WMP: Kankanaey *kulipangngáv* (\*kuli-) 'not to know which way to turn, where to go; take the wrong road, go astray', Kankanaey *atingawngáv* (\*qati-) 'not to know which way to turn, where to go; take the wrong road, go astray'

#### LOUD NOISE

WMP: Cebuano *alibángu* (\*qali-) 'bothering others by being noisy', Cebuano *alingása* (\*qali-) 'annoyingly noisy', Cebuano *alingísi* (\*qali-) 'making a piercing, screeching noise', Cebuano *alingisngis* (\*qali-) 'making very high-pitched and drawn out piercing noise', Cebuano *alingugngug* (\*qali-) 'noisy in a constant drumming way', Cebuano *alintabu* (\*qaliN-) 'turn over, producing a loud turmoil', Cebuano *salibagyaw* (\*sali-) 'annoyingly noisy; having the head awhirl, not knowing what to do; noise that is disturbing', Ifugaw *alibaddóng* (\*qali-) 'trampling and stamping impatiently', Ifugaw *alidogdóg* (\*qali-) 'droning noise made by those who pound rice', Ifugaw *alikadóng* (\*qali-) 'make a trampling noise, as children playing'

## DRUNK

WMP: Ifugaw *alingangá* (\*qali-) 'stunned, as an intoxicated man', Kankanaey *ali-bagbág* (\*qali-) 'be beside oneself for drunkenness'

Finally, several languages have \*qali/kali- words for the category 'conical':

## CONICAL

WMP: Bikol *balisúngsúng* (\*bali-) 'whirlpool, vortex; eddy; cone or funnel-shaped', Ilokano *balisongsóng* (\*bali-) 'render cone-shaped', Tagalog *alimulón* (\*qali-) 'conical, tapering', Hanunóo *salikungkúng* (\*sali-) 'funnel-shaped form, funnel-fashioned leaf, paper etc.'

Reconstruction: Proto Philippines \*balisurǰúŋ 'cone-shaped, funnel-shaped'.

All of the above notions appear to be connected, directly or indirectly, to the primary referents 'whirlwind/whirlpool', reflecting the shape (hair whorl, conical), the motion (sling, go round in circles), or the psychophysical effects of a human being whirling around (dizzy, confused, lost, drunk, with the abstract notion of confusion then extended back to other physical causes of sensory obfuscation such as thick smoke or the visually blurring effects of distance).

## 5 Residual categories

If we acknowledge that semantic contagion may have extended the range of application of the \*qali/kali- prefix by 'infecting' other categories through associations based on perceptual similitude, we can maintain that all \*qali/kali- words examined so far refer either to creepy-crawly creatures or to prodigies of Nature, together with extensions based on metaphor or shared abstract properties. What makes an understanding of the \*qali/kali- affix particularly challenging is that, despite the patterning seen so far, there are many other categories which appear to carry the same morphological marking.

One of the best-attested and semantically most transparent transitions from 'whirlwind/whirlpool' is that to 'hair whorl/crown of the head'. This is the first example seen of an \*qali/kali- word which refers to a part of the body. It is, however, not the only example. Others include: 'clavicle/collar bone', 'palate', 'pupil of the eye', and apparently 'scapula':

## CLAVICLE/COLLARBONE

F: Puyuma *Haliwazangan* (\*qaNi-) 'clavicle, collarbone'

WMP: Cebuano *balikhaw* (\*bali-; syncope?) 'collarbone; have prominent collarbones', Hanunóo *balískug* (\*bali-; syncope?) 'clavicle, collarbone', Toba Batak *haliadang* (\*kali-) 'clavicle, collarbone', Ilokano *aliwadáng* (\*qali-) 'collarbone, clavicle', Isneg *aliwadāng* (\*qali-) 'collarbone, clavicle', Kankanaey *alimadáng* (\*qali-) 'clavicle, collarbone', Kankanaey *aliwadáng* (\*qali-) 'rib (used only in tales)'

Reconstruction: PAN \*X-wadaŋ 'clavicle, collarbone'.

## PALATE

WMP: Kapampangan *alangálang* (\*qala-) 'palate, roof of the mouth', Iban *kelekanit* (\*kali-) 'roof of the mouth'

PUPIL OF THE EYE

F: Saisiat *ʔalimasawaʔan* (\*qaNi-) ‘pupil of the eye’

WMP: Tagalog *balintatáʔo*, *balintatáw* (\*baliN-) ‘pupil of the eye’ (cf. *táʔo* ‘human being, person’), Aklanon *kalimutáw* (\*kali-) ‘iris (of eye), eyeball’, Hiligaynon *kalimutáw* (\*kali-) ‘eyeball’, Bikol *kalintatáw* (\*kaliN-; cf. Tagalog *balintatáw* and Puyuma *muRTaTaw* ‘eyeball’) ‘centre of the eye containing the iris and the pupil’, Bikol *alinawnáw* (\*qali-) ‘centre of the eye containing the iris and the pupil’, Isneg *lintotólay* (\*qaliN-; cf. *tólay* ‘man, person’) ‘pupil of the eye’, Tagalog *alikmatá* (\*qali-) ‘pupil of the eye’, Bikol *alintatáw* (\*qaliN-) ‘centre of the eye containing the iris and the pupil’, Palauan *chelsúl a mad* (\*qani-?) ‘pupil of the eye’

Oc: Ponapean *limarepeileng* (\*qali-) ‘pupil of the eye (honorific)’

Reconstruction: PWMP \*X-Ca-Cau ‘pupil of the eye’ (lit. ‘person of the eye’).

SCAPULA

WMP: Angkola-Mandailing Batak *halipkip* (\*kali-) ‘scapula, shoulder bone’, Hanunóo *alipʔip* (\*qali-) ‘scapula, shoulder bone’, Kelabit *liʔip* (\*qali-; with syncope and cluster reduction?) ‘scapula’

As with terms for creepy-crawly creatures and natural phenomena, body-part terms can be arranged in marked and unmarked series.<sup>9</sup>

**Table 9:** Body-part terms, marked and unmarked with the \*qali/kali- prefix

UNMARKED	MARKED
1. hair (*bukeS)	hair whorl
2. rib (PMP *Rusuk, *tageRaŋ)	clavicle
3. gums (*gusi)	palate
4. tongue (PMP *dilaq)	
5. eye (*maCa)	pupil of eye
6. shoulder (*qabaRa)	scapula
7. bone (*CuqelaN)	

Table 9 aligns similar body parts in marked and unmarked categories: ‘hair’ is unmarked, while ‘hair whorl’ is marked’, ‘rib’ is unmarked, while ‘clavicle’ is marked, and so on. The number of unmarked body-part terms could be multiplied many times over, but this is unnecessary to make the point that the terms in the marked column are exceptional in their length, and in the phoneme sequences that form their onset syllables.

Other semantic categories for which \*qali/kali- words are fairly well-attested include names of birds, fish, and plants:

<sup>9</sup> Perhaps to be included here also is the bizarre set of apparent \*qali/kali- forms represented by Kankanaey *kalimputóy* (\*kaliN-) ‘calf, thigh – that is, the fleshiest part of them’, Tae’ *kalumpani*’ (\*kaluN-) ‘thin fold of flesh between the ribs and hind legs of a pig’, Ilokano *arimongmóng* (\*qari-) ‘the particles of fat distributed in the adipose tissue under the skin of the abdomen of swine’, Isneg *arimúyut* (\*qariN-) ‘hind part of an animal’s thigh’, Isneg *talínabáw* (\*sali/tali-) ‘the thigh of a hog that is given by the owner of a house where a solemn sacrifice took place, to his sister, as a present to take back home to her husband’, Kankanaey *alipadpád* (\*qali-) ‘side of the thigh – that is, the upper part’, Tae’ *tingkoran* (\*qatiN-) ‘thighbone of men and animals; thigh of the foreleg of a pig, commonly given in offering to the gods’.

## BIRDS (various)

WMP: Bolaang Mongondow *boyokuak* (\*bala-) 'kind of swamp bird', Hanunóo *balikáku* (\*bali-) 'medium-sized bird with long legs and reddish plumage, but otherwise presenting a parrot-like appearance', Hanunóo *balináyaw* (\*bali-) 'very small bird with black and white striped plumage', Hanunóo *balisúsu* (\*bali-) 'medium-sized red-billed kingfisher', Sangir *balindangeng* (\*baliN-) 'kind of sea bird', Maranao *kalasiansiang* (\*kala-) 'blue bird with white breast and long bill', Ngaju Dayak *kalialing* (\*kali-) 'small jet black bird with gray back', Sasak *kalidapang* (\*kali-) 'bird with red neck, black breast and wings, and red beak and legs', Tiruray *kelifodo?* (\*kali-) 'bird sp.', Tiruray *kelimetan* (\*kali-) 'hornbill sp.', Maranao *kalinsasaoi* (\*kaliN-) 'monkey bird', Hanunóo *kalusisi* (\*kalu-) 'very small hanging parakeet' Isneg *kulipagpāg* (\*kuli-) 'kind of bird', Aklanon *amaeádyang* (\*qama-) 'small black bird with red eyes'

## FIGHTING COCK (variety)

WMP: Cebuano *balakiki* (\*bala-) 'chicken coloured black with white speckles, and sometimes with other colours', Ilokano *boliála* (\*buli-) 'cock with yellowish plumage', Tae' *kaliabo* (\*kali-) 'chicken with black and brown feathers and black legs', Angkola-Mandailing Batak *hatinangke* (\*kati-) 'white fighting cock with black legs', Toba Batak *hatinangke* (\*kati-) 'white fighting cock with yellow legs', Aklanon *alimbúyog* (\*qaliN-) 'having many colours (such as certain chickens)', Bikol *alimbuyógon* (\*qaliN-) 'cock the colour of the black bee (*alimbubúyog*)', Ilokano *alimbuyógen* (\*qaliN-) 'cock with very dark red plumage', Tagalog *alimbuyugin* (\*qaliN-) 'cock with blackish feet and black spots on wings', Tagalog *talisayin* (\*tali-) 'green-spotted gray (said of roosters)'

Reconstruction: Proto Philippines \*qalimbuyugen 'cock with intense coloration (resembling that of a bumblebee)'

## OMEN DOVE

WMP: Toba Batak *darapati* (\*dara-) 'domestic dove', Nias (1) *kalafasi*, *kalafati* (\*kala-) 'domestic dove', Tagalog (1) *kalapati* (\*kala-) 'domestic dove', Karo Batak *kalibetah* (\*kali-) 'green dove with red throat and head', Kayan *kalibuken* (\*kali-) 'green imperial pigeon: *Ducula aenea*', Cebuano (2) *alimúkun* (\*qali-) 'kind of wild dove with white ears and light brown feathers speckled with black: *Phapitreron leucotis*', Ilokano (2) *alimúkeng* (\*qali-; -/ng/ irreg.) 'wild dove with gray plumage', Maranao (2) *limoken* (\*qali-) 'wild gray dove', Bolaang Mongondow (3) *limbukan* (\*qaliN-; /e/ irreg.) 'wild dove with sombre, mournful cry', Kadazan (3) *himbukon* (\*qaliN-) 'hill pigeon', Karo Batak (3) *limbukan* (\*qaliN-) 'dove sp.', Malay *lengguak*, *lengkuak* (\*qaliN-) 'thick-billed green pigeon: *Butreron capelli*', Malay (3) *limbok*, *limbukan* (\*qaliN-) 'pigeon generally ... but only in the language of sorcerers; bronzewing dove', Minangkabau *limpatu* (\*qaliN-) 'pigeon or dove, sp. unident.', Sangir *tarakuku* (\*tara-) 'turtle dove'

CMP: Buru *ermuken* (\*qari-) 'dove'

Reconstructions: 1. PWMP \*kala-pati 'domestic dove',<sup>10</sup> 2. Proto Philippines \*qali-muken, PMP \*X-muken 'omen dove', 3. PWMP \*qalim-buken 'omen dove'.

<sup>10</sup> Gonda (1973:165ff.) regards all quadrisyllabic variants of the word for 'domestic dove' with an apparent stem *-pati* as cases of "a Tamil element reaching the Archipelago in its Sanskritized form (*parapati*)". If this interpretation is correct the present comparison is a striking example of convergence both with regard to the phonetic modification of loanwords and with regard to the creation of apparent, but spurious, \*qali/kali-forms.

## OWL

WMP: Karo Batak *kalingkupa* (\*kaliN-) 'large owl sp.', Ilokano *kolalábang* (\*kula-) 'large own with gray plumage; feeds on chickens', Karo Batak *alingkupa* (\*qaliN-) 'large owl sp.'

Oc: Marshallese *lijemao* 'short-eared owl: *Asio flammeus*'

## SWALLOW

F: Saisiat *kaLkaLiliS* (\*kali-; syncope?) 'swallow'

WMP: Bare'e (1) *kalapini* (\*kala-) 'swallow sp. which nests on sheltered walls', Ilokano (1) *kalapini* (\*kala-) 'small bird with gray plumage; it lives near the water', Minangkabau *kalalatau* (\*kala-) 'kind of swallow', Taosug *kalasiyaw* (\*kala-) 'swallow', Ngaju Dayak *kalialang* (\*kali-) 'a small, very black bird with gray back', Karo Batak *kalimpini* (\*kaliN-) 'kind of swallow', Rungus Dusun *kalumpisau* (\*kaluN-) 'swallow sp.', Tae' *kalumpini*, *kaluppini* (\*kaluN-) 'swallow', Itbayaten *alpasayaw* (\*qali; syncope?) 'bird: *Apus pacificus*; *Hirundo tahitica jananica* (NM); swallow', Bolaang Mongondow *talimburung* (\*saliN/taliN-) 'swallow that makes edible nests', Sangir *salumpito* (\*saluN-) 'kind of swallow that makes edible nests'

CMP: Kambera *kalewaru* (\*kale-) 'kind of swallow: *Collocalia esculenta* D., Kambera *kaliwaru* (\*kali-) 'swallow, *Hirundo* D.'

Reconstructions: 1. PWMP \*kala-pini 'swallow sp.', 2. PWMP \*X-pinis 'swallow sp.'

## WOODPECKER

F: Saisiat (1) *baLasok* (\*bala-; haplology)

WMP: Cebuano (1) *balalatuk* (\*bala-), Malay (1) *belatuk* (\*bala-; haplology), Maranao (1) *balalatok*, Western Bukidnon Manobo (1) *tem-belelatuk* (\*bala-), Bare'e *walitut* (\*bali-), Bolaang Mongondow *bolingongo*, *olingongo* (\*bali-), Uma *balintutu?* (\*baliN-), Makasarese *bantinotto* (\*banti-), Isneg *kalitaxá* (\*kali-), Isneg *kalutaxá* (\*kalu-) 'woodpecker', Maranao *kolompia?* (\*kuluN-) 'woodpecker (white breast, red head, black back)', Iban *selematong*, *sematong* (\*sali-) 'spiderhunter, woodpecker'

Reconstruction: PAn \*bala-laCuk 'woodpecker'.

## FISH (various)

As noted already, some names of fish have been acquired by transfer from names of insects etc. through a perceived resemblance between the two, as with Cebuano *alibangbang* 'butterfly; butterfly fish: *Chaetodon* sp.; by extension, angel fish: *Holocanthus* sp.; kind of seashell'. Most, however, appear to be independent terms which acquired the \*qali/kali- affix for reasons inherent in the cultural valuation of these semantic categories themselves.

WMP: Makasarese *kalaus* (\*kala-) 'kind of smelt: *Sillago sihama*', Makasarese *kalampeto* (\*kalaN-) 'kind of edible sea fish', Makasarese *kalampute* (\*kalaN-) 'kind of edible speckled sea fish, about 20 cm. in length', Bolaang Mongondow *kolinama*, *kolindama* 'kind of marine fish', Casiguran Dumagat *kalibongbóng* (\*kali-) 'spadefish: *Scatophagus argus*', Maranao *kalinapad* (\*kali-) 'cyprinid in Lake Lanao', Tae' *kalussambang* (\*kalu-) 'kind of river fish', Ilokano *kurimaóng* (\*kuri-) 'kind of fish', Ilokano *kuritangtáng* (\*kuri-) 'edible marine fish', Cebuano *alásúus* (\*qala-) 'kind of fish: *Sillago* sp.', Cebuano *alimúsan* (\*qali-) 'kind of fish raised in fish ponds: *Paraplotosus albilabris*', Chamorro *alimasat* (\*qali-) 'kind of fish', Ilokano *alidengdeng* (\*qali-) 'small marine fish (bluish back, white belly)'



## MOLLUSK, SHELLFISH

WMP: Bare'e *balatani* (\*bala-) 'marine mollusk', Cebuano *balimbúgay* (\*baliN-) 'kind of bivalve', Bare'e *kalakapu* (\*kala-), Bare'e *kalancapu* (\*kalaN-) 'marine mollusk', Malay *kelembuai* (\*kaliN-) 'land snail of the genus *Ampulla*', Ilokano *kulintípay* (\*kuliN-) 'mollusk (shell used for window glass)', Cebuano *alakáak* (\*qala-) 'edible mollusk resembling the chiton, about an inch in length', Cebuano *alipadnu* (\*qali-) 'edible freshwater limpet', Malay *lengkitang* (\*qaliN-) 'snail, *Melania* spp.', Ilokano *arasés* (\*qara-) 'edible gastropod mollusk', Ilokano *ariesyés* (\*qari-) 'edible gastropod mollusk'

Oc: Marshallese *likaebeb* 'cone shell', *likajjid* 'money cowrie: Cypraeidae'

## PLANTS

A great many different plants are marked by reflexes of the \*qali/kali- prefix. The following is a brief selection:

F: Paiwan *qaLingelud* (\*qaNi-) 'a plant: *Liquidambar formosana*', Western Paiwan *quLimatsilaw* (\*quNi-) 'plant with small, grape-like fruit', Paiwan *quLimatsilu* (\*quNi-) 'a plant: *Ampelopsis heterophylla*', Paiwan *quLitsapudus* (\*quNi-) 'the paper mulberry: *Broussonetia papyrifera*'

WMP: Bare'e *balabati* (\*bala-) 'kind of pandanus', Hanunóo *balíknun* (\*bali-; syncope) 'tree sp.: *Melochia umbellata*', Hanunóo *balinána?* (\*bali-) 'flowering jungle vine having many thorns', Hanunóo *balináwnaw* (\*bali-) 'a tree, *Erioglossum ribiginosum*, Pinatubo Negritos *balinaknák* (\*bali-) 'woody vine: *Embelia* sp.' (Fox 1953:239), Bare'e *balincusu* (\*baliN-) 'tree with edible fruits', Hanunóo *balintawák* (\*baliN-) 'yellowish sweet potato having a light-coloured skin and black vine', Kankanaey *bulinaknák* (\*buli-) 'a tree (used only in tales)', Bare'e *kalamaya* (\*kala-) 'kind of creeping plant', Bare'e *kalamente* (\*kala-) 'plant sp.', Bare'e *kalantawu* (\*kalaN-) 'plant sp.', Bare'e *kalijawa* (\*kali-) 'plant sp.', Karo Batak *kaliméka* (\*kali-) 'edible mushroom', Karo Batak *kalisio* (\*kali-) 'tree with edible leaves', Malay *kelemayoh* (\*kali-) 'a brinjal, sp. unident.', Malay *kelepayang*, *kepayang* 'a tree: *Pangium edule*', Pinatubo Negrito (1) *kalibangbáng* 'a tree: *Bauhinia* spp.', Tagalog *kalimáyo* (\*kali-) 'local name for *kaláyo*: *Erioglossum ribiginosum*, a shrub with compact bushy crown', Karo Batak (1) *kalimbangbang* (\*kaliN-) 'tall, upright tree: *Alangium begoniifolium* Baill.', Karo Batak *kalincayo* (\*kaliN-) 'kind of small tree with leaves resembling rose leaves', Karo Batak *kalinjuhang* (\*kaliN-) 'the multipurpose plant *Cordyline fruticosa* Bakker (family Liliaceae)', Malay *kelempadang* (\*kaliN-) 'a shrub: *Vaccinium malaccense*', Malay *kelempayang* (\*kaliN-) 'a climber: *Pericampylus incanus*', Malay *kelempayan* (\*kaliN-) 'a plant name (variously identified)', Malay *kelempening* (\*kaliN-) 'a tree: *Pasania kunstleri*', Malay *kelempeti* (\*kaliN-) 'a tree: *Aporosa benthamiana*', *kelemunting* (\*kaliN-) 'rose myrtle', Toba Batak *halimbukbuk* (\*kaliN-) 'a shrub', Tagalog *kalumbibít* (\*kaluN-) '*Caesalpinia crista*, a prickly woody vine', Isneg *aladángan* (\*qala-) 'a low herb with small leaves that grows on tree trunks. Shamans take it along with them when they go to *maxanító* (communicate with the spirits)', Aklanon *alibútbut* (\*qali-) 'medicinal shrub: *Tabernaemontana pandacaqui*', Aklanon *alipáta?* (\*qali-) 'poisonous tree: *Excoecaria agallocha*', Ilokano (2) *alibangbáng* (\*qali-) 'a tree whose sour leaves are used for culinary purposes', Isneg *alipánay* (\*qali-) 'low herb that grows in forests', Malay *lemesu* (\*qali-) 'a shrub: *Matthaea sancta*', Simalur (2) *alifambang* (\*qali-) 'kind of tree', Tagalog (2) *alibambáng* (\*qali-) 'small stocky tree with leaves used for flavouring meat or fish: *Bauhinia malabarica*', Isneg *alimbanógan* (\*qaliN-) 'a forest tree with white flowers', Isneg *alimbató* (\*qaliN-) 'large forest tree with nut-like fruits', Isneg *alimboboxó* (\*qaliN-) '*Paspalum* sp. A kind of grass: crush its leaves, boil them and rub the body with them in order to cure the



*kudilaw* itch', Isneg *alimbódo* (\*qaliN-) 'herb that is placed near the spot where one begins to clear the ground for a new rice field: it protects the farmer from any kind of harm', Isneg *alimbuxáy* (\*qaliN-) 'a tree whose small, round red fruits are threaded on a piece of string to be worn around the head, in order to cure fever and diseases of the eye', Malay *lembayong* (\*qaliN-) 'the water hyacinth: *Eichornia crassipes*', Malay *lembéga* (\*qaliN-) 'a plant: *Calotropis gigantea*', Malay *lembesu* (\*qaliN-) 'a tree: *Fagraea* spp.', Malay *lembiding* (\*qaliN-) 'a fern: *Tenochlaena palustris*', *lembugai* (\*qaliN-) 'horse-radish: *Moringa pterygosperma*', Malay *lempaung* (\*qaliN-) 'tree yielding an edible sour fruit', Malay *lempenai* (\*qaliN-) 'tree sp.', Malay *lempoyan* (\*qaliN-) 'a plant: *Stereospermum fimbriatum*', Malay *lempoyang* (\*qaliN-) 'a ginger (*Zingiber aromaticum* or *Z. zerumbet*) used medicinally', Malay *lempunai* (\*qaliN-) 'a tree: *Xylopiia caudata*', Malay *lemputeh* (\*qaliN-) 'small tree: *Urophyllum griffithianum*', Malay *lenggada* (\*qaliN-) 'a tree (used medicinally): *Diospyros lucida*', Malay *lenggadai* (\*qaliN-) 'a mangrove: *Bruguiera parviflora*', Malay *lenggapus* (\*qaliN-) 'a tree: *Mesua ferrea*', Tagalog *alinsánay* (\*qaliN-) 'wild banana', Tagalog *alintátaw* (\*qaliN-) 'medium-sized tree the bark of which is used in decoctions for coughs: *Diospyros pilosanthera*', Aklanon *aeógbáti* (\*qalu-) 'a vegetable: *Basella rubra* L.', Aklanon *aeopísan* (\*qalu-) 'a vine: *Tetrastigma harmandii*', Aklanon *aeosíman* (\*qalu-) 'herb that grows along bank of river, purslane: *Portulaca oleracea*', Cebuano *salibutbut* (\*sali-) 'a shrub the leaves and sap of which have medicinal uses: *Tabernaemontana pandacaqui*'

Reconstructions: 1. PWMP \*kali-bañbañ 'a tree, probably *Bauhinia* spp.', 2. PWMP \*qali-bañbañ 'a tree, probably *Bauhinia* spp.'

Given the wide range of plants represented in the above (very restricted) list, one might ask whether it is meaningful to cite plants as a group as evidence for the \*qali/kali- prefix. Words for 'nettle', however, suggest that the atypical canonical shape of many other plant words is no accident:

#### NETTLE

F: Puyuma *ringaten* (\*qali-) 'tree nettle: *Laportea pterostigma*', Paiwan *qaLaLipetj* (\*qaNa-) 'nuisance plant: *Cyanotis kawakamii*', Puyuma *lingadaRan* (\*qaNi-) 'nettle (*Urtica thunbergiana*)'

WMP: Isneg *alalátāng* (\*qala-) 'a dioecious, urticaceous shrub with very irritating, stinging hairs and large leaves', Hanunóo *alingangát* (\*qali-) 'a nettle-like plant', *alingátung* (\*qali-) 'lipa (*Laportea meyniana* Warb.?), a nettle-like plant with stinging hairs on the underside of the leaves', Malay *linggata* (\*qaliN-) 'a nettle-like plant, unident.'

Finally, various \*qali/kali- semantic categories are encountered which have no obvious connection to others. In the interest of brevity these will not be documented here, but include the following: dandruff; rash/sores (hence combined as 'skin disease'); expand upward; feline quadruped (PWMP \*qari-maun); gargle, rinse the mouth (PMP \*qali-muRmuR, with some languages reflecting just \*muRmuR, or other affixed forms); mote in the eye; numb; red clouds of sunset; shy/timid; squirrel (= flying squirrel?); talk/walk in one's sleep; topsy-turvy (PWMP \*balin-tuaj).

We have now surveyed nearly the full range of meanings marked by the \*qali/kali- prefix in Austronesian languages. In accordance with more general historical tendencies toward conservative or innovative morphology, this affix is best preserved in the Formosan languages, the Philippine languages and some of the languages of western Indonesia, as the Batak languages of northern Sumatra. Other languages of western Indonesia, as Iban or

Malay, preserve many examples, but because of regular vowel neutralisations in prepenultimate syllables have often lost information about the vocalic part of the affix variants retained. In general, the languages of eastern Indonesia and those in the Oceanic group preserve only vestiges of the original system.

The suggestion that \*qali/kali- affixation once constituted a system which is largely fossilised in the modern languages confronts us again with the issue of meaning or function. What, if any, is the common thread that runs through the semantic categories we have now examined? The honest answer to this question probably is 'none'. Categories such as 'sling' or 'conical' almost certainly are historically secondary, the result of 'semantic contagion', arising from more basic categories such as 'whirlwind/whirlpool'. Semantic contagion presumably was a recurrent historical process, and shared innovations should exhibit a dendritic pattern. However, some likely extensions, as 'hair whorl', evidently were present in Proto Austronesian itself, suggesting that the process of expanding the categorial representation of \*qali/kali- forms had already begun before the break-up of Proto Austronesian. Do these complications vitiate further attempts to understand this bizarre yet widespread feature of Austronesian word structure, or is there some coherent pattern that underlies the surface complexity?

## 6 The spirit world

Traditional religious beliefs in all their variety are commonly grouped under the unifying term 'animism', a term first proposed by Tylor in 1871 (Tylor 1958), and richly illustrated within the context of a theory of the spirit world which maintains much of its validity today. Just as the major world religions are defined by a shared body of belief motivated by similar ideology, so is animism. What is different about the two is that the major world religions have acquired their attested distributions through diffusion, often in the form of conquest or forced adoption. By contrast, animism is widely shared as a result of convergent psychological adaptations to the common problems of coping with the stresses and mysteries of the material world. In other words, each of the major world religions began in a single centre or with a single individual and spread through indoctrination, while the ideas of animism arose independently again and again in the minds of many unconnected human beings. For this reason it makes some sense to speak of animism as the 'natural' religion of humanity. More than any body of inculcated doctrine the common properties of animistic beliefs must reflect universal properties of human psychology. If they did not their very universality would constitute one of the great mysteries of science.

Like others of his generation, Tylor was a comparativist rather than a descriptivist. As such, he sought useful generalisations about all human cultures, and in nearly all cases he was forced to rely upon data provided by travellers or missionaries rather than by trained professional observers. When descriptive anthropology or ethnography began to develop in the twentieth century the work of Tylor and others of his time came under attack in part on the grounds that it sought to compare decontextualised culture traits. How could beliefs which are embedded in different overall cultural contexts possibly have the same meaning for participants in those differing cultures? And if they do not have the same meaning in different cultures how can they possibly be treated as equivalent for comparative purposes? At one time such criticisms were taken as damning, but even a little reflection shows that the criticisms themselves are open to serious objections.

Nineteenth-century anthropology followed much the same course as nineteenth-century linguistics in that comparative theory (historical linguistics; ethnology) raced ahead of descriptive theory (descriptive linguistics; ethnography). If it is methodologically inadmissible to compare decontextualised culture traits then it should be equally inadmissible to compare decontextualised linguistic traits, since linguistic facts are debatably just a particular type of culture trait. But it would be ludicrous to argue that English *eye* cannot be compared with its German cognate *Auge* on the grounds that each term is embedded in a different linguistic system which gives it a unique and therefore noncomparable meaning. As historical linguists have known since early in the nineteenth century, it is impossible to compare whole systems: the **only** practical basis for comparison is the morpheme. Much the same is true of typological comparison as it has developed in the twentieth century. No one would consider trying to compare whole systems; rather, one compares word-order typology, types of possessive marking, relativisation, or the like, which of necessity are extracted from the larger context in which they function within the system of a language.

The preceding remarks are critical in laying out the theoretical basis for the interpretations which follow. Since animistic beliefs tend strongly toward shared basic universal properties regardless of differences in detail, evidence for the association of particular facets of nature with the spirit world need not be drawn from the same societies for which linguistic data are cited. All that matters is to show, for example, that butterflies—which are marked with \*qali/kali- morphology in many Austronesian languages—are **universally** regarded as visible signs of departed spirits. Whether or not this culture universal is overtly manifested in a society which uses an \*qali/kali- word for ‘butterfly’ is irrelevant. Both cultural and linguistic change are constant and inevitable, and cultural change often precedes linguistic change, leaving linguistic relics as evidence of a once more-highly integrated past. What we appeal to is not the synchronic correlation of linguistic form and religious belief in an attested culture, but the marking of semantic categories by the \*qali/kali- prefix in Austronesian languages and evidence of a universal association between that category and the world of spirits. In the absence of overt evidence that a particular animistic belief is present in a given culture, then, we accept **covert** evidence of its psychological presence. Covert evidence is manifested by the distribution type: distributions which cannot plausibly be attributed to inheritance from a historically inferrable common ancestor or to borrowing must be due to the independent operation of psychological tendencies that are pan-human.

If we make allowance for semantic contagion, a fairly strong case can be made that many of the lexical categories marked by \*qali/kali- share an important common property, although it is neither a linguistic property, nor a semantic property which can be perceived in the natural world. Rather, what defines many \*qali/kali- words, and distinguishes them from unmarked lexical categories of similar semantic content, is a dangerous connection with the world of spirits. Raw material for documenting this connection is given in the Appendix, and will be discussed only briefly here. To begin with ‘creepy-crawly creatures’, butterflies are linguistically marked as the visible forms of ghosts in Kayan of central Borneo, where *hiap toʔ* ‘butterfly’ literally means ‘ghost chicken’, and in Malagasy, where *lolo* ‘butterfly; ghost’ is polysemous in an unusually revealing way. In Isneg of northern Luzon *kulibangbáng* means ‘butterfly, moth’, but “In prayers, it often stands for ‘spirit’” (Vanoverbergh 1972:323). While this trait distribution probably would be sufficient in itself to establish that similar unreported associations of ‘butterfly’ and ‘spirit’ are found in many other Austronesian-speaking societies (and may well have been found in their common ancestor), statements from dictionaries or ethnographies confirm the universal character of this

association, which is reported also among Sino-Tibetan and Austroasiatic-speaking peoples (see Appendix). Similarly, leeches are included among those creatures supernaturally protected in the widely distributed 'thunder complex' of insular Southeast Asia and the western Pacific (Blust 1981, 1991).

The next category, 'ant', reveals another characteristic of \*qali/kali- morphology. Many Austronesian languages make multiple lexical distinctions for different species of ants, in contradistinction to 'butterfly, moth', which is typically marked by a single term. The generic term 'ant' (PMP \*sejem) carries no \*qali/kali- marking, but particular ant species do. Much the same appears to be true of other creepy-crawly creatures for which Austronesian languages typically recognise several lexical distinctions, as caterpillars, crabs, snakes and spiders (Table 6).

In the next general category, prodigies of Nature, the rainbow has powerful connections with the spirit world, contrasting sharply with such 'ordinary' natural phenomena as rain. In traditional animistic societies the rainbow is most commonly represented as an enormous snake which drinks the rain. As the manifestation of a frightening spiritual presence it must be accorded due respect, hence the globally distributed taboo against pointing at the rainbow with the index finger (Blust n.d.). Similarly, whirlwinds and whirlpools, echos, parhelia, sunshowers and the like all have significant dangerous connections with the world of spirits in many traditional societies globally.

Two objections can be raised against this interpretation of the data. First, it cannot be shown that all categories which carry \*qali/kali- marking have such a cultural association. What dangerous connection with the spirit world can the scapula, for example, or gargling possibly have? Second, in animism spirit beings are ubiquitous: if virtually *anything* can harbor a minatory spirit presence, isn't the motive force behind use of the \*qali/kali- prefix lost?

As an answer to the first objection it seems reasonably clear that some lexical categories which carry \*qali/kali- marking have acquired this morphological status through 'semantic contagion'. We would thus not necessarily expect categories such as 'sling', 'conical', or 'sparks' to have dangerous associations with the spirit world (although they might for other reasons). Moreover, animistic beliefs are not among the best-documented aspects of most traditional cultures, often being mentioned only in passing. In some cases it is likely that gaps in the documentary record conceal cultural connections between \*qali/kali- words and dangerous spirit associations which would become apparent with fuller ethnographic documentation.

As for the second objection, it is true that animistic belief systems often acknowledge a natural order that is permeated with spirit presences, but it would be a mistake to view animism as a kind of pretheistic pantheism. Some cultures may very well have significant animistic associations in connection with categories such as 'housefly' or 'rain', but it is abundantly clear on a global basis that such associations are far more salient and important in categories such as 'butterfly' and 'rainbow'. In short, spiritual forces tend to reside in exceptional natural objects or events (hence the opposition of marked and unmarked lexical categories in Tables 6, 8 and 9).

Finally, if further evidence is needed, names of spirits appear with extremely high frequency as \*qali/kali- words. This is particularly true in those more marginal societies in which the system of animistic beliefs was not transformed through exposure to Islam or Christianity. In a dictionary remarkable for its cultural sensitivity the Belgian priest Morice Vanoverbergh recorded no fewer than thirty-three Isneg names of spirits which are

quadrisyllabic forms beginning with reflexes of the now familiar \*qali/kali- onset (*ala-, ale-, ali-, aliN-, bala-, balaN-, bali-, baliN-, bara-, bula-, buli-, buri-, kala-, kali-, kula-, tala-, tali-, taliN-, talo-*). The following is a brief selection which could be extended considerably:

## SPIRIT

WMP: Isneg *balasingnúd* (\*bala-) 'a spirit who lives in the *síxay* hut', Isneg *balikádan* (\*bali-) 'a spirit whose name is pronounced by shamans when in their trances', Isneg *balikúdan* (\*bali-) 'another name for the *agbalikádan* spirit', Isneg *balinawāng* (\*bali-) 'a helpful spirit', Isneg *balingatáy* (\*bali-) 'a man-killing spirit who strikes his victims at the neck', Isneg *balingató* (\*bali-) 'a spirit with the same habit as the preceding', Isneg *balintawāg* (\*baliN-) 'a female spirit who ... lives in a pool of the Apayaw river, just below Sabangan. She never emerges, but nobody should go there', Cebuano (1) *bulalákaw* (\*bula-) 'harmful supernatural being that takes the form of a ball of fire, with trailing sparks. If it brushes or gets close enough to smell the skin, it makes a permanent white spot', Kankanaey (1) *bulalákaw* (\*bula-) 'kind of animal (?), supposed to be an old eel, to fly and to be luminous', Western Bukidnon Manobo (1) *Bulelakaw* (\*bula-) 'spirit deity of stream and lakes', Isneg *burinána?* (\*buri-) 'a spirit who causes a general swelling of the body, more especially of the abdomen', Isneg *kalapátaw* (\*kala-) 'a spirit who rules the entire ornithic kingdom', Kapampangan *kaladuwá* 'soul' (cf. *duwa* 'two', and Tagalog *kaluluwá* 'soul, spirit, vital principal', the latter probably with /ka/- plus reduplication), Ngaju Dayak *kalabawai* (\*kala-) 'forest spirit', Isneg *kalibutág* (\*kali-) 'female spirit who lives in the water and uses one of her hairs to ensnare the person she wants to drown', Casiguran Dumagat *kaliduwa* (\*kali-) 'soul, spirit of a living person', Iban *kelemayang* (\*kali-) 'dim figure, shadow, ghost', Maranao *kalilangan* (\*kali-) 'altar where evil spirits are appeased with sacrifices', Bolaang Mongondow *kalintuang* (\*kaliN-) 'spirit that walks on head and hands', Malay *kelembayan, lemboyan* (\*kaliN-) 'invisible elves of the forest and shore', Kankanaey *aladunáxan* (\*qala-) 'spirit who is wont to make little children cry so as to disturb the parents in their sleep', Kankanaey *alaláyo* (\*qala-) 'name of a spirit used in prayers', Isneg *alibowá* (\*qali-) 'a spirit', Isneg *alimangáw* (\*qali-) 'a spirit who is sometimes visible, but whether seen or not, frightens people so that their hair stands on end', Kankanaey *aliliá* (\*qali-) 'phantom, spectre, ghost, spirit', Kankanaey *alingáwan* (\*qali-) 'spirit's child', Kankanaey *alipungdān* (\*qali-) 'malicious spirit', Kankanaey *aliwáwa?* (\*qali-) 'spirit who roams in and around the village, and is never able to stop walking', Maranao *alimekat* (\*qali-) 'spirit, god of the water', Maranao *linibeng* (\*qali-) 'god of the unseen beings', Isneg *alimbabakóng* (\*qaliN-) 'spirit who lives in roomy houses', Ilokano *aningáas* (\*qani-) 'kind of ghost', Ilokano *aniwáas* (\*qani-) 'ghost', Timugon Murut *timbunus* (\*qatiN-) 'spirit of a comet (can terrorise pregnant women)' (Prentice 1971:64), Isneg *talimúngāt* (\*sali/tali-) 'a spirit who ... inflicts sickness on any person who passes in the vicinity of his abode', Isneg *talipagdóxān* (\*sali/tali-) 'a spirit who ... has bees for pets. If a man hunting for bees sees those that belong to that spirit, he becomes sick', Isneg *talipáso* (\*sali/tali-) 'a spirit who ... never sleeps', Iban *Selempandai* (\*saliN-) 'deity, creator of matter (incl. iron) and maker of man', Iban *Selempatah* (\*saliN-) 'creator of men', Isneg *talimbukawtabúkaw* (\*saliN/taliN-) 'a spirit, the husband of *talimbukitabúkit*', Isneg *talimbukitabúkit* (\*saliN/taliN-) 'a female spirit, the wife of *talimbukawtabúkaw*', Maranao *salindagao* (\*saliN-) 'evil spirit'

Oc: Ponapean *likamisik* (\*qali-) 'horrible in appearance; anything peculiarly sacred'

Reconstruction: Proto Philippines \*bulalakaw 'kind of nature spirit, probably identified with the appearance of a comet'.

However, we have yet to make a case for the link between cultural belief and linguistic marking: why would a dangerous connection with the spirit world require that a special affix be added to the relevant lexical categories? As should be evident by now, many of the referents of nominal \*qali/kali- forms are of a type likely to be associated with taboo. In fact, the names of several taboos or physical symptoms commonly associated with violation of a taboo are marked with \*qali/kali-:

#### TABOOS

WMP: Bare'e *kalanoa* (\*kala-) 'swelling of genitals' (from breaking a taboo?), Tae' *kaliuanan* (\*kali-) 'violate a taboo', Toba Batak *halispison* (\*kali-) 'get a swollen foreskin as a result of urinating into the hearth fire',<sup>11</sup> Isneg *karibosót* (\*kari-) 'abnormal inflation of the abdomen' (from breaking a taboo?)

Taboos, like any part of culture, must be learned, and given their importance to the adult members of a society it probably is critical that children not be allowed to ignore or to flout them beyond the minimum which can be expected during the process of enculturation. In an incisive critique of Edmund Leach's well-known essay 'Anthropological aspects of language: animal categories and verbal abuse', Halverson (1976:508) has observed that "if an object is taboo, it must, for reasons of avoidance, etc., be recognized with greater clarity than other objects, not less". An effective means of facilitating such recognition—particularly in societies which frown upon corporal punishment of children, as do Austronesian-speaking societies generally—would be through the use of distinctive linguistic marking. It is precisely children rather than adults who are likely for lack of other preoccupations to step on ants, stretch earthworms till they break, crush centipedes or millipedes under sticks or stones, pull the wings from termites, beetles or butterflies, or point impulsively at the sudden, seemingly miraculous appearance of a rainbow. How could a child's attention be drawn to those categories of experience which require culturally circumscribed behaviour even in the absence of adult supervision? If Proto Austronesian and its early descendants lacked quadrisyllabic stems, as present evidence indicates, \*qali/kali- words would have had high perceptual salience, thereby drawing immediate attention to the cultural sensitivity of the lexical category so marked, and hence facilitating children's acquisition of critical portions of their culture. Given the rich morphological system that must be attributed to Proto Austronesian and many of its descendants, it is true that many other words (but not stems) would have been quadrisyllabic. But the great majority of these were verbs or deverbal nouns which would have been quite distinct from \*qali/kali- words.

As the Appendix should demonstrate, this interpretation goes some way toward explaining why many \*qali/kali- words mark the semantic categories they do, and not others. Moreover, the hypothesis that \*qali/kali- marked semantic categories which have a dangerous connection with the spirit world has two ancillary benefits. First, it provides a natural explanation for the generally fossilised character of the affix in \*qali/kali- words, since the function of such an affix would have required that it be more-or-less **obligatory**. Second, although many questions remain unanswered, this hypothesis suggests a reason for hyperallomorphy. The force of most bound morphemes is carried by their phonemic content, not by their canonical consequences for the affixed word. By contrast, \*qali/kali- lengthened disyllabic bases into

<sup>11</sup> The taboo against urinating into a hearth fire is widespread, with similar supernatural consequences in a number of An-speaking societies, including the Thao of central Taiwan, and in Fiji (Paul Geraghty, pers. comm.).



canonically distinctive quadrisyllabic words. As a result, the phonemic content of the affix was less important than the number of syllables it contained, and consequently was free to vary. Statistical frequency shows that \*qali- and \*kali- are the most frequently encountered forms of the affix, but not necessarily the earliest shapes. Subgrouping considerations support \*qaNi- and, to a lesser extent, \*qali- as prototypical \*qali/kali- variants, but lend little support to the antiquity of \*kali-, which is rare in the Formosan languages.

In short, then, the function of the \*qali/kali- prefix evidently was to mark facets of experience that were regarded as spiritually dangerous, hence requiring special precautions of a sort likely to be violated by incompletely acculturated children. It did this purely by lengthening the affixed word to an atypical quadrisyllabic shape, hence marking the associated semantic categories as those requiring particular behavioural sensitivity.<sup>12</sup>

## 7 Postscript

While this paper was being edited a serendipitous discovery was made. Ken Rehg recalled that during the weekly Micronesian seminars held in the Department of Linguistics at the University of Hawai'i from 1976 to 1981 Byron dropped in one day to announce that he had noticed something interesting in Marshallese which had previously escaped his attention. Marshallese, he said, has what appear to be fossilised prefixes, *li-* and *la-* in many names of plants, insects and other types of animals, and also as feminine and masculine markers respectively. This observation probably grew out of his work on the Marshallese dictionary which he co-authored (Abo et al. 1976). Byron's announcement evidently produced a flurry of activity among the members of the seminar,<sup>13</sup> as Ken discovered (to his surprise) a long-forgotten folder labelled 'li/la prefixes', which contained: (i) a three-page list of Carolinian *li-* forms having to do with (a) people, (b) birds, (c) insects, (d) plants, (e) shells, (f) fish, crabs and sea creatures, and (g) miscellaneous; (ii) a three-page list of Marshallese words with initial *li-* and *la-* formatives having to do with (a) shells, (b) other small creatures, (c) weapons, (d) games and pastimes, (e) topography and elements, (f) stars, constellations and legendary figures, (g) body parts, (h) dishes, (i) actions and qualities, (j) plant names, (k)

<sup>12</sup> Based on a much narrower body of data, two writers have suggested a more restricted function for the \*qali/kali- prefix. In his grammar of Timugon Murut (Sabah) Prentice (1971:118) identified a morpheme *liN-* with allomorphs *li-* and *liN-* which reportedly occurs 'in names of plants and animals'. He specifically mentions names for 'scorpion', 'water leech', 'intestinal worm', and a plant (*Costus speciosus*). The most ambitious attempt to deal with \*qali/kali- words is that of Kähler (1949–50), who suggested that many names of trees, plants, and animals (especially fish and birds) in the languages of Indonesia and the Philippines contain fossilised 'classifiers'. Thus, *tulang* 'bone' is reported as a formative for tree names in various languages (Bikol *tulang nanok*, Malay *tulang daing*), and is assumed to underlie more oblique derivations, as Minangkabau *lagundi* 'kind of shrub' (said to be from \**tulang gundi*). Trisyllabic tree names that begin with *ka-* or *ke-* are said to derive from reduced compounds with \**kayu* 'tree', while others with *ta-* or *taN-* reflect \**batang* 'trunk', as with Samoan *tamanu* 'a tree: *Calophyllum inophyllum*' (said to be from \**batang* 'trunk' plus a purely hypothetical base). A second set of disyllabic formatives listed as (1) *mara-*, *mala-*, *maja-*, *moro-*, *molo-*, and (2) *kali-*, *hali-*, *ali-*, *bali-*, *koli-*, *oli-* is said to occur in plant and animal names, in terms for illnesses, and only rarely in concrete nouns of other kinds. The animals so marked are characterised collectively as (1) unpleasant, (2) appearing in swarms, and/or (3) quick.

<sup>13</sup> These are identified elsewhere (Bender & Wang 1985:83) as "Byron W. Bender, Robert W. Hsu, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel and Judith Wang" who formed the core group, as well as Paul Geraghty, Ward H. Goodenough, Sheldon P. Harrison and a number of graduate students who were less regularly associated with it.

fish, (l) canoe parts, and (m) miscellaneous; (iii) a one-page list of Woleaian *la-* and *li-* formants having to do with (a) fish, (b) shells, (c) birds, (d) other creatures, (e) games and (f) miscellaneous; (iv) eight Pulo Annian and two Sonsorolese words containing an apparent *ni-* formative and referring to a fruit dove, crab, spider, jelly fish, bachelor/spinster, a fish (the perch), twins and doll; (v) several dozen Mokilese words which appear to contain *la-* and *li-* formatives referring to (a) fish, (b) birds, (c) female persons, (d) various other creatures and (e) an assortment of other semantic categories; (vi) a scattering of Kosraean words containing *IV-* and referring to (a) men's names, (b) women's names, (c) actions and qualities, (d) creatures, (e) games and weapons, (f) a food and (g) topography and elements; and (vii) a three-page list of Gilbertese (Kiribati) words containing an apparent formative *ni-* which refer to names of (a) fish, (b) shellfish and other marine life, (c) plants, (d) insects, (e) birds, (f) games, (g) body parts, (h) ailments, (i) star names, constellations and (j) water.

No dates appear on any of this material, which never reached the publication stage, and individual responsibility for the lists is not indicated. Despite this vagueness in the historical record we can be fairly certain that it was Byron who compiled the Marshallese list and who stimulated other members of the seminar to search their own language for similar phenomena. As can be seen, the semantic categories identified in the Micronesian material only partially overlap those found in the more broadly representative Austronesian material of the present study. In particular, the correlation of \**li-* with feminine categories and \**la-* with masculine categories is not known to have parallels anywhere outside Micronesia, and may represent a historically independent phenomenon which in Micronesian languages has become confounded with reflexes of \**qali/kali-*. Nonetheless, it is of particular interest to see that some twenty years ago Byron discerned in the Marshallese material a fossilised affix or affixes which may reflect, at least in part, the \**qali/kali-* prefixes documented here.

## Appendix

This appendix presents a brief ethnological survey of the semantic categories connected with the \**qali/kali-* prefix. Its major purpose is to show, by way of primary references, that the semantic categories marked by the \**qali/kali-* prefix in Austronesian languages are commonly associated with the world of spirits. A word of explanation is in order.

The claim made here is that the types of belief reported in this survey reflect universals of animism. Culture universals, like linguistic universals, are elements which need not be present in all sample units, but which occur with a distribution that implies independent development, and hence some general pragmatic, psychological or behavioural motivation. Universals of animism are either overtly represented in the belief system of a culture, or are latent in that system. Latency can be viewed as a nonchance tendency for some feature to appear over time, and so is Janus-faced: it is likely that universals which are not present in a documented belief system were present in an ancestral form of that system, or will appear in some future form of the system. For this reason it is not necessary to establish exact correlations between the presence of an \**qali/kali-* affix in a language and a supporting belief about the spirit world in the same language community. Rather, it is sufficient to establish that a universal of animism exists, and could have motivated the innovation of the \**qali/kali-* prefix. To illustrate, the word for 'rainbow' is marked with this affix in thirteen languages. Ethnological data supporting the claim that the rainbow is commonly viewed as a powerful spiritual presence is available for a number of cultures, but for the most part these do not correspond in the available data to the thirteen languages which carry the distinctive linguistic marking. This lack of correspondence is regarded as irrelevant to the argument, since:



(i) there are many gaps in the ethnological data regarding traditional belief systems, and (ii) the firm establishment of a culture universal leaves no choice but to conclude that the cognitive or behavioural details which manifest that universal are likely to have been actualised in the past. A great deal of relevant material on animism in Indonesia is to be found in Kruijt (1906), but given this type of argument it follows that ethnological data offered to establish universals of animism need not be restricted to Austronesian-speaking societies, and some supporting evidence is accordingly drawn from other ethnolinguistic groups where appropriate. Non-Austronesian-speaking groups are marked with a double asterisk.

ANT. Iban (Sarawak): *sampok* "termites, Isoptera, ants ... Mounds signify fertility and are left unharmed; even under houses, where the floor may be cut away to allow growth." (Richards 1981:321)

Isneg (northern Luzon): *aliw(a)liwāt* "Middle-sized, very black, stinging ant. It is supposed to resent the presence of coconut oil: should a person staying in the forest boil coconut milk in order to extract the oil, that ant will change the guilty person into a *kāngaw*, a fabulous wild animal." (Vanoverbergh 1972:354). *Latón* "the nest of termites. Climbing it causes the abdomen to swell." (Vanoverbergh 1972:48)

Malay: "Nests, either of a large species of black ant or of the termite, are sometimes thought to be the dwelling places of spirits." (Evans 1923:269)

Manobo (Mindanao): "Big biting ants are believed to be the spirit of a dead relative who feels lonely." (Demetrio y Radaza 1970:1:183)

Negritos of Rizal (Luzon): "In Rizal ... ant hills were supposed to be inhabited by spirits known as *matánda sa punsu*, that is, 'old man of the ant hill'. Wherever I went hills were inhabited by such old men. The believers in them always asked permission to take wood or whatever they needed if it was in the vicinity of the ant hill." (Garvan 1964:226-27)

Sa'a (Southeast Solomons): "At Sa'a, and in the neighboring parts of Malanta, the same word is used for the soul of a living man and the ghost of an ordinary person, '*akalo*.'" After death "the mere '*akalo* soon turn into white ants' nests, which again become the food of the still vigorous ghosts; hence a living man says to his idle son 'When I die I shall have ants' nests to eat, but what will you have?'" (Codrington 1891:260)

AUREOLE. \*\*Lakher (Sino-Tibetan, Assam): "Parhelia are known as sawmachupa; they are very unlucky, and are believed to portend an unnatural death." (Parry 1932:500)

\*\*Scots: "In the far north of Scotland, parhelia are regarded as ill-omened and as forerunners of bad weather." (Parry 1932:500, fn. 1)

BEE. Among the Malays of the Malay peninsula a punitive thunderstorm can be provoked by blowing on the nest of a kind of small bee. (Evans 1937:178ff.)

BUTTERFLY. Isneg (northern Luzon): "*kulibangbāng*. Butterfly, moth. In prayers it often stands for 'spirit'." (Vanoverbergh 1972:323)

Kayan (central Borneo): *heñap tu?* 'butterfly' (lit. 'spirit/ghost chicken') (Blust 1977:99)

Malagasy: *lolo* 'butterfly, moth; ghost' (Richardson 1885:398)

Nias (Barrier Islands, west of Sumatra): "We also know of the people of Nias that they regard the souls of their dead (when these have left no son behind) as turned into butterflies." (Kruijt 1906:175)

Sundanese (west Java): Children are told not to catch butterflies in the cupped hands. If they do so their hands will be *hileud-eun*, that is, supernaturally scorched—just as the index finger is, if one points at a rainbow. (Dudu Prawiraatmaja, pers. comm.)

Yami (Imorod dialect; Botel Tobago island, southeast Taiwan): *pahapahad no anitu* 'butterfly' (= unknown element + genitive marker + 'ghost') (Tsuchida et al. 1987:175)

\*\*Nagas (Sino-Tibetan, Assam): "Some believe that the soul takes the form of various insects, especially butterflies." (Nag 1964:49)

\*\*Pulau Tawar (Austroasiatic, Malay peninsula): "The spirits of the dead become white butterflies and it is therefore tabu to kill these insects." (Evans 1923:210)

DRAGONFLY. Negritos of Malaya (Austroasiatic): "Violations of the commandments of the Thunder God are reported to him by divine messengers which are conceived as dragonflies or wasps." (Blust 1981:299)

EARTHWORM. Toraja (central Sulawesi): "The Torajas regard it (the earthworm) as the fugitive soul-stuff of a person." (Kruijt 1906:180)

ECHO. \*\*English: "The Anglo-Saxon dictionary preserves the curious word *woodmare* for an echo (*wudu-mær* = wood nymph), a record of the time when Englishmen believed, as barbarians do still, that the echo is the voice of an answering spirit." (Tylor 1960:213)

\*\*Hausa (Afroasiatic, northern Nigeria): "The echo is attributed to a supernatural agency, in fact it is called *Tblis*, devil, or *Kurua*, meaning soul, spirit, shadow." (Tremearne 1913:112)

Malays: "*hantu*. Evil spirit; ghost ... They may be grouped as follows: (i) demons of localities ... (ii) demons tied to special spots or tutelary spirits of freaks of nature ... (iii) demons behind natural phenomena, such as: echoes (*hantu-hantu-an*)." (Wilkinson 1959:395)

Puluwat (central Carolines, Micronesia): *Yi ya kkapah, yeray hoomá ya likáhenwan ááy kkapah* 'I spoke, a ghost echoed my speech' (Elbert 1972:66)

FIREFLY. With its eerie bioluminescence it is not hard to see why the firefly is associated with the world of spirits in many cultures worldwide. Data from Isneg, Karo Batak, Kwaio, Sa'a and Toraja suggest that there was an early An belief in the danger of fireflies entering a house. The specific basis of this belief is further clarified by the form of the belief as it is reported for Isneg and Toraja: if a firefly should enter a house and perish in the hearthfire it was a portent of human deaths to follow. In some incompletely understood way it appears likely that the sparks from a fire were seen as one manifestation of this sacred insect. For further details on the connection of fireflies with the spirit world in Indonesia see Kruijt (1906:171ff.).

Buru (central Moluccas): "In Buru if one sees a firefly near a grave it is believed that the *nitu* (soul) of the deceased has come to visit his resting place." (Kruijt 1906:172)

Gitua (Rai coast, New Guinea): *mate* 'die', *yap* 'fire', *mate yap* 'firefly' (Lincoln 1977:17)

Hanunó (Mindoro, central Philippines): *buyinaw* 'firefly, believed to be a bringer of bad luck' (Conklin 1953:97)

Isneg (northern Luzon): *alipatpat* "firefly or glow-worm. If it enters the house and comes into collision with fire, a person dies for every spark that it produces." (Vanoverbergh 1972:47)

Karo Batak (northern Sumatra): "*kalimpetpet*. Firefly; a bad omen if it enters a house: thieves will come." (Neumann 1951:134)

Kwaio (Malaita, southeast Solomons): "*bulubulu*. Star; firefly, believed (especially when it enters house) to be messenger from *adalo* (ancestral spirit, ghost)." (Keesing 1975:29)

Numfor (Cenderawasih Bay, Irian Jaya): "*naser*. Phosphorescent creatures, both fireflies and marine organisms. The fireflies that are known as *naser* are of a somewhat larger sort

than the *manimanjar* ... The *naser* and not the *manimanjar* are said to be manifestations of the spirits of the dead.” (van Hasselt and van Hasselt 1947:164)

Rennellese (southeast Solomons): “*agito* ... Firefly, *Luciola* sp., Wolff ... seen inland, a rare phenomenon; they were believed to be the embodiment of the goddess Sikingimoemoe and might not be killed.” (Elbert 1975:7)

Sa'a (Malaita, southeast Solomons): “Fireflies are popularly regarded as ghosts and are killed by the children if they enter the house.” (Ivens 1927:353)

Sundanese (west Java): “According to the Sundanese fireflies are ghosts that trek over mountain and field bearing burning torches.” (Kruijt 1906:173)

Tehit (West Papuan; Irian Jaya): Fireflies are *sétan* or *lampu sétan* (evil spirits/lights of evil spirits) and are dangerous to people. (Don A.L. Flassy, pers. comm.)

Toraja (central Sulawesi): “One evening as a Toraja chief was sitting a firefly entered his house, and he strove vigorously to keep the insect away from the resinous torch inside. When asked why he did this he answered ‘when a firefly burns one of us will die’.” (Kruijt 1906:171)

Toraja: “Whenever a swarm of fireflies settles on a tree near a house the Torajas think that someone in the house will soon die, because the fireflies are the souls of the dead (according to others, their eyes) who have come to take a victim away.” (Kruijt 1906:172)

Tuaran Dusun (north Borneo): “The firefly (*nenekput*) is the spirit of a dead man.” (Evans 1923:16)

GARGLE/RINSE THE MOUTH. Ethnographic evidence of a connection between gargling and the spirit world remains elusive. However, it is noteworthy that plants which carry \*qali/kali-morphology may be important as mouthwashes among some Austronesian-speaking groups:

Pinatubo Negritos (west-central Luzon): “The roots of the following plants are boiled in water and the decoction used as a mouthwash for toothaches: ... *alipungpung*: *Borreira articularis* Linn.” (Fox 1953:341)

GRASSHOPPER. Isneg *asisít* “1. a small grasshopper with red eyes that comes to people who call it by means of a small clicking sound of the tongue. Children should abstain from doing so, as it might kill them by removing their *kaduduwā* (soul), 2. a man-killing spirit who slays his victims through the eyes.” (Vanoverbergh 1972:96–97)

Rennellese (Polynesian Outlier, southeast Solomons): '*atua segesege ba'e* '1. harmful supernatural or person. Lit., leg-cutting supernatural, 2. a large grasshopper' (Elbert 1975:21)

LEECH. Leeches are among those creatures guarded by supernatural sanctions in the widespread ‘thunder complex’ of Australasia (Needham 1964; Blust 1981, 1991). Needham (1964:141ff.) reports a belief among both the Negritos of the Malay Peninsula and the nomadic Penan of Borneo that leeches must not be burned or it will precipitate the wrath of the thunder god, causing the perpetrator to be struck by lightning and turned to stone. While Needham believes that it is the burning of blood rather than the burning of the leech itself that is forbidden, the fact that leeches are just one of many creepy-crawly creatures marked with the \*qali/kali- prefix (and so labelled as objects of taboo) suggests that it is blood, rather than the leech that is contingent in this taboo.

MILLIPEDE. Although millipedes in general may very well have connections with the spirit world among some Austronesian-speaking peoples, it appears to be significant that most recorded \*qali/kali- words for ‘millipede’ refer to phosphorescent varieties. Note Malinowski’s remark that the Kiriwina sorcerer is “feared as ghosts are feared by us, as an uncanny manifestation. One is afraid of meeting him in the dark, not so much because he

might do any harm, but because his appearance is dreadful and because he has at his bidding all sorts of powers and faculties which are denied to those not versed in black magic. His sweat glows ..." (Malinowski 1961:421) The following details therefore refer to bioluminescence more generally. Needless to say, a connection with 'firefly/glow-worm' through this common property is likely.

\*\*Jehehr (Malayan Negritos): "It is tabu for anyone to kill a millipede, to shoot a certain species of owl with the blowpipe, or to flash a looking glass or other shining object, about in the open." (Evans 1923:153)

Motu (Central District, Papua New Guinea): *mamaro* 'a phosphorescent insect on surface of sea, supposed to be spirit of a dead man' (Lister-Turner & Clark 1930:102)

Tanga (New Ireland): *ka:nu* 'phosphorescent light emanating from lichens and regarded as a manifestation of the evil spirit Tara' (Bell 1977:35)

OMEN DOVE. Bagobo (Mindanao): "A common method used by the spirits to communicate with mortals is through the call of the *limokon* (a dove, *Calcophops indica*). All the people know the meaning of its calls and all respect its warnings." (Cole 1913:108)

Among the Bagobo "It is taboo for a youth who has never killed a man to eat the flesh of the *limokun* pigeon" (which has a sacred association with omens). (Benedict 1916:241)

Malay: Among Malays the flesh of the turtle dove is forbidden. (Skeat 1900:190)

Pinatubo Negritos (west-central Luzon): "It seems that the souls or spirits of demised infants fly around in the form of a certain variety of wild dove common in these parts." (Garvan 1964:118)

Subanun (western Mindanao): *limukun* "A bird of evil omen; when seen or heard it postpones work." (Finley & Churchill 1913:196)

Ulawa (southeast Solomons): "At Ulawa the crested pigeon *toowao*, and a cuckoo, both of which are rare birds, are regarded as omens of sickness when they appear and cry." (Ivens 1927:353)

OWL. No doubt in part because of their nocturnal habits and large disquieting eyes, often seen on dark nights before any outline of a perching body is visible, owls are commonly associated with the dead in many cultures.

\*\*Ainu (isolate; Japan): "There are five specific birds whose cry should not be imitated by anyone. They are the cuckoo, woodpecker, nighthawk, goatsucker and owl. These birds have power to bewitch people by means of their cry, and sometimes they do." (Batchelor 1901:409)

Balinese *cahak* 'owl, believed to bring bad luck' (Barber 1979:65)

\*\*Esselen (Hokan?; central California coast): "According to Navarrete (1802), the Esselen believed that they were transformed into owls at death." (Hester 1978:498)

Ifugaw (northern Luzon): "*tukukun* = owl = one of the *lubug* (creatures and objects that give omen)." (Barton 1955:234, fn. 6)

\*\*Jehehr (Malayan Negritos): "It is tabu for anyone to kill a millipede, to shoot a certain species of owl with the blowpipe, or to flash a looking glass or other shining object, about in the open." (Evans 1923:153)

Malay: *burung hantu* 'owl' (lit. 'ghost bird')

Puyuma: *HalTu* 'owl' (from PAN \*qaNiCu 'ghost, spirit of the dead')

PLANTS. For the connection of many plants with the spirit world see Kruijt (1906:136ff.), who pays particular attention to the cordyline or dracaena, and Fox (1953:305ff.).

**PUPIL OF THE EYE.** The magico-religious import of the pupil is not entirely clear. Tylor (1958:15) points out that the pupil of the eye is widely associated in European folklore with personal animation. Barnes (1974:107ff.) relates a Kédang myth in which a civet cat was raised by orphans. When it died the pupils of its eyes were buried in front of the door of their house, and these grew into a magical tree of wealth.

**RAINBOW.** Buli (south Halmahera): "Rai. Rainbow. According to the Buli people the rainbow possesses a supernatural force; one may not point at it or his finger will wither." (Maan 1940:92)

Cebuano Bisayan (central Philippines): "If one points at a rainbow his index finger will be supernaturally cut off." (Martinianna van Dierendonck, pers. comm.)

Dusun (northern Borneo): Among the Tuaran Dusun of Sabah "It is forbidden, or rather it is unwise, to point at the rainbow, as the finger that you use to point with will rot away." (Evans 1923:15)

Isneg (northern Luzon): "*bunglún*. The rainbow. Supposed to be a spirit ... Persons who point at it with their index finger (not: another finger) will be troubled, on the following night, with either whitlow or the destruction of the tip of the index or the loss of its nail ... *mabunglún*. A swelling of the abdomen caused by drinking water that has been touched by a rainbow." (Vanoverbergh 1972:182)

Javanese: In the village of Keji, central Java "There used to be a belief ... that to point at a rainbow would cause a bent finger ... To counter the effect you should poke the finger in buffalo dung." (S.O. Robson, pers. comm.)

Kankanaey (northern Luzon): "*Ngális di áso*. A superstitious formula, pronounced when seeing a rainbow, lest it eat one's soul and so one become thin." (Vanoverbergh 1933:317)

Kédang (Lembata, Lesser Sundas): The people of Kédang fear "pointing at a *nado-tado*, that is a rainbow or any other rising effulgence of a spirit. If one does point one runs the risk that one's finger will be permanently bent." (Barnes 1974:216)

Sundanese (west Java): Children are warned not to point at a rainbow lest the index finger be *hileud-eun* (supernaturally scorched). (Dudu Prawiraatmaja, pers. comm.)

**RED CLOUDS OF SUNSET.** Universally associated with blood, war and spiritual danger.

Bagobo (Mindanao): "When the western sky has a lurid or reddish aspect on a cloudy afternoon, it is a sign of misfortune for the world, and it especially foretells the appearance of the sickness called *pamalii*." (Benedict 1916:245)

Malay: "Sunset is the hour when evil spirits of all kinds have most power. In Perak children are often called indoors at this time to save them from unseen dangers." (Skeat 1900:15)

Manobo (Mindanao): "At sunset it is not good for young children to be still playing on the ground, because there are evil spirits who could pass by, and if you bump them you will be sick." (Demetrio y Radaza 1970:3:473)

Tehit (West Papuan, Irian Jaya): "The golden glow of evening is dangerous, and everyone rushes indoors to avoid it." (Don A.L. Flassy, pers. comm.)

Trukese (Eastern Carolines, Micronesia): *Wúúmeeyòn* "Spirit of the late afternoon ... It makes infants sick; therefore women traditionally avoided carrying their infants outside of the house in the late afternoon." (Goodenough and Sugita 1980:389)

**RESTLESS.** Restless movements as seen in unquiet sleep, the death struggle of fish out of water, or birds being slaughtered may recall the erratic flight of the butterfly.

Philippines (region unspecified): "The concept of the *kalagkalag*, or the restless dead, who sometimes 'bother' the living relatives by sending them sickness of undiagnosed origin is still very much with the lowland peoples." (Demetrio y Radaza 1970, 1:vii)

SHADOW/REFLECTION. In animistic thinking the shadow or reflection is all but universally regarded as a visible manifestation of the soul, the same word often serving to mark both meanings.

Isneg (northern Luzon): "*aniníwing*. Shadow, reflection. A child who looks at its shadow in the evening becomes very thin." (Vanoverbergh 1972:69)

SKIN DISEASE. Frazer (1960:548) comments that "the eating of a sacred animal is often believed to produce leprosy or other skin disease". In Austronesian-speaking societies, at least, this statement can be generalised to the breaking of a taboo of almost any kind. Skin diseases are commonly ascribed to spirit attacks.

Cebuano (central Philippines): skin ailments frequently are said to be "the result of the influence of spirits, with water as the medium of contagion". (Richard W. Lieban, pers. comm.)

Ilokano (northern Luzon): *kaybaán* 'a fairy supposed to inhabit nests of white ants or coconut groves, and to help farmers, weavers, housekeepers etc., or to inflict skin diseases etc.' (Carro 1956:140)

Isneg (northern Luzon): *rasā*. "The itch or scabies. Causes by mites. A person may get this skin disease by touching fireflies at night or by being covered by the bristles of the *baxināt* (*Saccharum* sp.)." (Vanoverbergh 1972:472)

Pinatubo Negritos (west central Luzon): "The Negritos do not normally explain skin diseases as being caused by the spirits, rather it is the common belief that dermatosis is inherited, but if any sickness is serious, the spirits are always involved." (Fox 1953:321)

SNAKE. Although use of the \*qali/kali- prefix suggests that only certain types of snakes have associations with the spirit world, Kruijt (1906:178) makes a more general claim: "In general the peoples of Indonesia usually see a snake as the incarnation of a dead person".

Bahau (central Borneo): "The soul-stuff of men once appeared in the form of a snake." (Kruijt 1906:178)

Nias (Barrier Islands, west of Sumatra): "The people of Nias say that the souls of insolent men turn into snakes after their death." (Kruijt 1906:178)

SUMMIT. Balantak (eastern Sulawesi): "mountain tops are considered sacred". (LeBar 1972:138)

SUNSHOWER. Arosi (San Cristobal, southeast Solomons): *arito* 'a sunshower; to shine in rainy weather; to clear up, of the weather (people keep indoors at such times for fear of ghosts)'

Javanese: A sunshower is a sign that someone important (king, high official) has died. (Soebardi, pers. comm.)

\*\*Kintak Bong (Austroasiatic, Malay peninsula): "Expectant women must not go out during 'hot rain' (i.e. rain when the sun is shining), which is much feared." (Evans 1937)

Kiribati/Gilbertese (Micronesia): *riringa ni moan atu* 'morning sun between two showers; sign of death' (Sabatier 1971:321)

Long Terawan Berawan *akang biloh kijih* 'the giant ghost who appears during periods of sun with rain' (Proctor 1979)

Malay: "Mid-day, when a light rain is falling and the sun is shining at the same time is usually regarded as equally dangerous to the golden glow of sunset." (Skeat 1900:15, fn. 4)

Manggarai (Flores, Lesser Sundas): *usang ta'a hitu, usang mélu, usang ata ba beti* "Sunshower, 'hot rain'. Such a rain brings multifarious illnesses." (Verheijen 1967:727)

TALK/WALK IN ONE'S SLEEP. This category hardly requires documentation, since the appearance of speech and motor actions in a human body which does not appear to be under its own control would readily be attributed to spirit possession in most traditional societies.

TOPSY TURVY. In many cultures the inversion of objects or reversal of sequences in ritual performances (as in the so-called 'Black Mass') has the effect of establishing a curse. One example, of possible European origin, is given here. A second associates inversion with incest, itself an act surrounded by the strongest taboos and punitive sanctions of a supernatural character.

\*\*Chorti (Mayan; Guatemala): "If the mere wishing of harm does not suffice, the Indian may burn a candle on the doorstep of the pueblo church at midnight, either on a Thursday or a Friday. It is stood upside down, with the wick end pointing toward the Devil, who is asked to send the harm to the victim." (Wisdom 1940:333)

Kédang (Lembata, Lesser Sundas): "One of the phrases for incest, *hunéq-koloq* means 'to turn upside down', is also applied to putting a house post in the ground in the reverse position from which it grew." (Barnes 1974:68)

WASP. \*\*Negritos of Malaya (Austroasiatic): "Violations of the commandments of the Thunder God are reported to him by divine messengers which are conceived as dragonflies or wasps." (Blust 1981:299)

WHIRLWIND/WHIRLPOOL. \*\*Ainu (isolate; Japan): "The Ainu imagine whirlwinds to be filled with demons, and they therefore fear them." (Batchelor 1901:385)

\*\*Chorti (Mayan; Guatemala): "Lagarto is a giant lizard spirit, probably of Spanish origin, which lives in the deep spots of streams and is the spirit of whirlpools ... He has a large tail, the end of which is of bone and shaped like an ax. His mouth is extremely large, so that he swallows his victims whole. People who come to the spot to bathe, especially at night, are liable to be killed with the axlike tail and swallowed." (Wisdom 1940:409)

\*\*Gabrielino (Uto-Aztecan, southern California): "Whirlwinds were evil spirits." (Bean & Smith 1978:548)

\*\*Germans: "The Loreley is only a modernised version of the river-demon who drowns the swimmer in the whirlpool." (Tylor 1960:213)

Isneg "*alipugpúg*. Whirlwind, eddy of air. It announces the presence of the spirit of the same name." (Vanoverbergh 1972:47)

\*\*Lakota/Dakota (Siouan; Dakotas): "The Dakota believe that there is a close relation between the whirlwind and the fluttering wings of a moth. The cocoon is regarded as the bundle or mysterious object from which a power similar to that of the whirlwind emanates ... In the whirlwind somehow and somewhere resides the power to produce confusion of mind ... When a man loses his presence of mind he is said to have been overcome by the power of the whirlwind." (Wissler 1905)

Maya (Mayan; northwestern Guatemala): "As in other parts of Mesoamerica, there is the belief in the evil eye, in the whirlwind as a dangerous supernatural, and in sickness caused by *susto* (fright)." (Wagley 1969:63)

Merina (Madagascar): "*tadió* ... A whirlwind. It is generally supposed to consist of the spirits of the dead." (Richardson 1885:597)



\*\*Tarahumara (Uto-Aztecan; northern Mexico): The Tarahumara regard the wind “as a good being, the whirlwind as an evil one. Whirlpool-beings are evil, fat and piglike, and can cause disease (seize souls).” (Fried 1969:863)

\*\*Tehit (West Papuan; Irian Jaya): Whirlpools are said to be the residence of an evil spirit. (Don A.L. Flassy, pers. comm.)

\*\*Toba (Guaicuruan; Paraguay): “Whirlwinds ... are regarded as the passing of the spirits. The Tobas are in the habit of saying of such whirlwinds: ‘There goes a *péyak* (= spirit) dancing in the dust’.” (Karsten 1932:120)

\*\*Yuman tribes of the Gila River (southern Arizona): “An individual possessed a plurality of souls”, one of which was “the whirlwind soul, the ghost”, which stays about (in human guise) as the whirlwind (Spier 1933:296)

WOODPECKER. \*\*Karen (Sino-Tibetan; peninsular Burma): The woodpecker is a bird of ill omen. (Marshall 1922:190, 229)

\*\*Malayan Negritos (Austroasiatic): The woodpecker is regarded as a sacred bird. (Evans 1923:182).

Subanun (Mindanao): *bulatúk* ‘a spirit bird that determines the best site for a house’ (Finley & Churchill 1913:185; cf. PAN \*bala-laCuk ‘woodpecker’)

## References

In the interest of brevity standard sources for language material are not given here, but many of these may be found in Blust (1980). Nonobvious sources are cited directly in the text. I am indebted to the late D.J. Prentice for making available material on Timugon Murut from a manuscript dictionary which remained unpublished at his death.

Abinal, P. and V. Malzac, 1963, *Dictionnaire malgache-français*. Paris: Éditions maritimes et d’Outre-mer. (First published in 1888.)

Abo, Takaji, Byron W. Bender, Alfred Capelle, and Tony DeBrum, 1976, *Marshallese-English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawai‘i.

Anceaux, J.C., 1961, *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde. The Hague: Nijhoff.

Barber, C.C., 1979, *Dictionary of Balinese-English*, 2 vols. Aberdeen, Scotland: Aberdeen University Library Occasional Publications, No. 2.

Barnes, R.H., 1974, *Kédang: a study of the collective thought of an eastern Indonesian people*. Oxford: Clarendon.

Barton, Roy F., 1955, The mythology of the Ifugaos. *American Folklore Society Memoirs* 46:1-244.

Batchelor, John, 1901, *The Ainu and their folklore*. London.

Bean, Lowell John and Charles R. Smith, 1978, “Gabrielino”. In Robert F. Heizer, ed. *Handbook of North American Indians*, vol. 8: California, 538-549. Washington: Smithsonian Institution.



- Bell, F.L.S., 1977, *Tanga-English, English-Tanga dictionary*. Oceania Linguistic Monographs, No. 21. University of Sydney.
- Bender, Byron W. and Judith W. Wang, 1985, The status of Proto-Micronesian. In Andrew Pawley and Lois Carrington, eds *Austronesian Linguistics at the 15th Pacific Science Congress*, 53-92. Canberra: Pacific Linguistics.
- Benedict, Laura Watson, 1916, A study of Bagobo ceremonial, magic and myth. *Annals of the New York Academy of Science* 25:1-308.
- Blust, Robert, 1977, Sketches of the phonology and morphology of Bornean languages, 1: Uma Juman (Kayan). *Papers in Borneo and Western Austronesian linguistics*, No. 2. Canberra: Pacific Linguistics.
- 1980, Austronesian etymologies. *Oceanic Linguistics* 19:1-181.
- 1981, Linguistic evidence for some early Austronesian taboos. *American Anthropologist* 83:285-319.
- 1991, On the limits of the "thunder complex" in Australasia. *Anthropos* 86.4/6:517-528.
- n.d., Pointing, rainbows and the archaeology of mind. MS., 134pp.
- Bradshaw, Joel, 1978, The development of an extra series of obstruents in Numbami. *Oceanic Linguistics* 17:39-76.
- Carro, Andrés, 1956, *Iloko-English dictionary*. Translated, augmented and revised by Morice Vanoverbergh. Baguio: Catholic School Press. (First published in 1888.)
- Chrétien, C. Douglas, 1965, The statistical structure of the Proto-Austronesian morph. *Lingua* 14:243-270.
- Codrington, R.H., 1891, *The Melanesians: studies in their anthropology and folklore*. Oxford: Clarendon.
- Cole, Fay Cooper, 1913, *The wild tribes of the Davao district, Mindanao*. Field Museum of Natural History publication 170, Anthropological Series, vol. 12, No. 2. Chicago.
- Conklin, Harold C., 1953, *Hanunóo-English vocabulary*. Berkeley and Los Angeles: University of California Press.
- Demetrio y Radaza, Francisco R., ed. and comp., 1970, *Dictionary of Philippine folk beliefs and customs*. 4 volumes. Cagayan de Oro City: Xavier University.
- Dempwolff, Otto, 1938, Vergleichende Lautlehre des austronesischen Wortschatzes, vol. 3. *Zeitschrift für Eingeborenen-Sprachen*, supplement 19. Berlin: Reimer.
- Elbert, Samuel H., 1972, *Puluwat dictionary*. Canberra: Pacific Linguistics.
- 1975, *Dictionary of the language of Rennell and Bellona*, 2 parts. Copenhagen: National Museum of Denmark.
- Evans, Ivor H.N., 1923, *Studies in religion, folk-lore and custom in British North Borneo and the Malay Peninsula*. Cambridge: Cambridge University Press.
- 1937, *The Negritos of Malaya*. Cambridge: Cambridge University Press.
- Ferrell, Raleigh, 1969, *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica, Monograph 17.

- 1982, *Paiwan dictionary*. Canberra: Pacific Linguistics.
- Finley, J.P. and W. Churchill, 1913, *The Subanu*. Carnegie Institution of Washington Publications 184:1–236.
- Fox, Robert B., 1953, The Pinatubo Negritos: their useful plants and material culture. *Philippine Journal of Science* 81:173–414.
- Frazer, Sir James George, 1960, *The golden bough: a study in magic and religion*, 1 vol., (abridged edn). New York: Macmillan. (First published in 1922.)
- Fried, Jacob, 1969, “The Tarahumara”. In Evon Z. Vogt, ed. *Handbook of Middle American Indians*, vol. 8: Ethnology, part 2:846–870. Austin: The University of Texas Press.
- Garvan, John M., 1964, The Negritos of the Philippines. In Hermann Hochegger, ed. *Wiener Beiträge zur Kulturgeschichte und Linguistik*, vol. 14. Veröffentlichungen des Instituts für Völkerkunde der Universität Wien.
- Gericke, J.F.C. and T. Roorda, 1901, *Javaansch–Nederlandsch Handwoordenboek*, 2 vols, (4th edn). Amsterdam and Leiden: Johannes Muller and E.J. Brill. (First published in 1847.)
- Gonda, Jan, 1952, The comparative method as applied to Indonesian languages. *Lingua* 1:86–101.
- 1973, *Sanskrit in Indonesia*, (2nd edn). New Delhi: International Academy of Indian Culture.
- Goodenough, Ward H. and Hiroshi Sugita, 1980, *Trukese–English dictionary*. Memoirs of the American Philosophical Society, vol. 141. Philadelphia: The American Philosophical Society.
- Guerreiro, Antonio, Vocabulary of Punan Kelai. MS., 5pp.
- Halverson, John, 1976, Animal categories and terms of abuse. *Man* 2 (n.s.), 505–516.
- Hasselt, J.L. van and F.J.F. van Hasselt, 1947, *Noemfoorsch woordenboek*. Amsterdam: de Bussy.
- Hester, Thomas Roy, 1978, “Esselen”. In Robert F. Heizer, ed. *Handbook of North American Indians*, vol. 8: California, 496–499. Washington: Smithsonian Institution.
- Ho, Dah-an, 1978, A preliminary comparative study of five Paiwan dialects. *Bulletin of the Institute of History and Philology*, Academia Sinica, vol. 49, part 4:565–681.
- Ivens, W.G., 1927, *Melanesians of the Southeast Solomon Islands*. London: Kegan Paul, Trench, Trubner.
- Jeng, Heng-hsiung, 1971, *A Bunun–English dictionary*. Taipei. Privately circulated.
- Kähler, Hans, 1949–50, Untersuchung über die Entstehung klassifikatorischer Präfixe in austronesischen Sprachen. *Zeitschrift für Eingeborenen-Sprachen* 35:162–191.
- 1961, *Simalur–Deutsches Wörterbuch mit Deutsch–Simaluresischem Wörterverzeichnis*. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg, vol. 3. Berlin: Reimer.

- Karsten, Rafael, 1932, *Indian tribes of the Argentine and Bolivian Chaco*. Ethnological Studies. Societas Scientiarum Fennica, Commentationes Humanarum Litterarum, vol. 4, No. 1. Helsinki.
- Keesing, Roger M., 1975, *Kwaio dictionary*. Canberra: Pacific Linguistics.
- Kruijt, Alb. C., 1906, *Het animisme in den Indischen Archipel*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde. The Hague: Nijhoff.
- LeBar, Frank M., 1972, "Balantak". In Frank M. LeBar, ed. *Ethnic groups of insular Southeast Asia*, vol. 1: *Indonesia, Andaman Islands and Madagascar*, 138–139. New Haven: Human Relations Area Files.
- Li, Paul Jen-kuei, 1977, The internal relationships of Rukai. *Bulletin of the Institute of History and Philology*, Academia Sinica, 48(1):1–92.
- 1981, Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology*, Academia Sinica, 52(2):235–301.
- Lincoln, Peter C., 1977, *Gitua–English vocabulary*. Typescript, 48pp.
- Lister-Turner, R. and J.B. Clark, 1930, *A dictionary of the Motu language of Papua*. Sydney: Government Printing Office.
- Maan, G., 1940, *Boelisch–Nederlandsch woordenlijst*. Bandoeng: Nix.
- McFarland, Curtis D., 1977, *Northern Philippine linguistic geography*. Study of Languages and Cultures of Asia and Africa monograph series, No. 9. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Malinowski, Bronislaw, 1961, *Argonauts of the western Pacific*. New York: E.P. Dutton. (First published in 1922.)
- Marshall, Harry Ignatius, 1922, *The Karen people of Burma: a study in anthropology and ethnology*. The Ohio State University Bulletin, vol. 26, No. 13, Contributions in History and Political Science, No. 8. Columbus.
- Nag, Moni, 1964, "Nagas". In Frank M. LeBar, Gerald C. Hickey and John K. Musgrave, eds *Ethnic groups of mainland Southeast Asia*, 44–49. New Haven: Human Relations Area Files.
- Needham, Rodney, 1964, Blood, thunder, and mockery of animals. *Sociologus* (n.s.), 14:136–149.
- Neumann, J.H., 1951, *Karo Bataks–Nederlands woordenboek*. Medan, Sumatra: Varekamp.
- Panitia penyusun kamus Bali-Indonesia (Committee for the compilation of a Balinese–Indonesian dictionary), 1978, *Kamus Bali-Indonesia*. Denpasar, Bali.
- Parry, Neville Edward, 1932, *The Lakhers*. London: Macmillan.
- Pratt, George, 1984, *A grammar and dictionary of the Samoan language*. With English and Samoan vocabulary. (3rd, rev. edn). Papakura, New Zealand: R. McMillan. (First published in 1893.)
- Prentice, D.J., 1971, *The Murut languages of Sabah*. Canberra: Pacific Linguistics.

- Proctor, John, 1979, A preliminary look at the language of Long Terawan. *The Sarawak Museum Journal* 27.48 (n.s.), 103–170.
- Reid, Lawrence A., ed., 1971, *Philippine minor languages: word lists and phonologies*. Oceanic Linguistics Special Publication, No. 8. Honolulu: University of Hawai'i Press.
- Richards, Anthony, 1981, *An Iban–English dictionary*. Oxford: Clarendon.
- Richardson, J., 1885, *A new Malagasy–English dictionary*. Antananarivo: London Missionary Society.
- Sabatier, E., 1971, *Gilbertese–English dictionary*. Translated from the French by Sister Oliva of the Catholic Mission, Tarawa. Tarawa: Sacred Heart Mission.
- Skeat, Walter William, 1900, *Malay magic: an introduction to the folklore and popular religion of the Malay peninsula*. London: Macmillan.
- Spier, Leslie, 1933, *Yuman tribes of the Gila River*. Chicago: The University of Chicago Press.
- Tremearne, A.J.N., 1913, *Hausa superstitions and customs*. London: John Bale, Sons and Danielsson.
- Tryon, D.T., 1976, *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.
- Tsuchida, Shigeru, 1976, *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa monograph series, No. 5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tsuchida, Shigeru et al., 1987, *Lists of selected words of Batanic languages*. Department of Linguistics, Faculty of Letters, University of Tokyo.
- Tylor, Edward Burnett, 1958, *Primitive culture*, 2 parts. New York: Harper and Brothers. (First published in 1871.)
- 1960, *Anthropology*. Ann Arbor: The University of Michigan Press. (First published in 1881.)
- Vanoverbergh, Morice, 1933, *A dictionary of Lepanto Igorot or Kankanay*. Anthropos Linguistic Monograph, vol. 12. St Gabriel-Mödling: Anthropos Institute.
- 1972, *Isneg–English vocabulary*. Oceanic Linguistics Special Publication, No. 11. Honolulu: The University Press of Hawai'i.
- Veen, H. van der, 1940, *Tae' (Zuid-Toradjasch)–Nederlandsch woordenboek, met register Nederlandsch–Tae'*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde. The Hague: Nijhoff.
- Verheijen, Jilis A.J., 1967, *Kamus Manggarai 1: Manggarai–Indonesia*. The Hague: Nijhoff.
- Wagley, Charles, 1969, The Maya of Northwestern Guatemala. In Robert Wauchope, ed. *Handbook of Middle American Indians*, vol. 7, *Ethnology*, part 1:46–68. Austin: University of Texas Press.
- Wallace, Alfred Russell, 1962 [1869], *The Malay archipelago*. London: Macmillan.

- Wilkinson, R.J., 1959, *A Malay-English dictionary* (romanised), 2 parts. London: Macmillan.
- Wisdom, Charles, 1940, *The Chorti Indians of Guatemala*. Chicago: The University of Chicago Press.
- Wissler, Clark, 1905, The whirlwind and the elk in the mythology of the Dakota. *Journal of American Folklore* 18:257-268.
- Zeitoun, Elizabeth, n.d., Word lists of Formosan languages. MS., 77pp.



# *The elusive shape of realis/irrealis in Jabêm*

---

JOEL BRADSHAW

A paradigmatic distinction between realis and irrealis mode pervades the verbal system of Jabêm, yet neither the morphological means by which it is marked nor the syntactic position in which it shows up is always reliable. The distinction is neutralised on more than two-thirds of verb stems and throughout the whole plural set of subject prefixes, so that almost half the verbs in running text remain unmarked for either realis or irrealis. Fortunately, the presence of verb serialisation in Jabêm ensures that over two-thirds of clauses contain at least one verb in which the distinction is overtly marked, although that marking is more likely to occur on a secondary verb than on the main verb in each clause. The elusive shape of this basic grammatical distinction may indicate that Jabêm is on the verge of reinforcing and eventually replacing its realis/irrealis morphology with an invariant predicate-initial particle, as several neighbouring languages have done.

## 1 Introduction

Jabêm (also spelled Yabem) is among the best-recorded languages in Melanesia. Dempwolff (1939) provides a coherent and linguistically sophisticated description of its major features, and Zahn (1940) provides a thorough pedagogical grammar, with abundant exemplification. Zahn earlier produced a (1917) Jabêm–German dictionary, and Streicher has produced both a (1937) German–Jabêm dictionary and a (1982) Jabêm–English dictionary. Since Jabêm was a church and mission-school lingua franca used since the turn of the century throughout Austronesian-speaking areas of Morobe Province, Papua New Guinea, there are also abundant materials in the language: primers, texts about a variety of traditional and modern customs and scientific beliefs, traditional Jabêm and biblical tales, hymnals and liturgy, the entire New Testament, and most of the Old Testament (Streicher 1982:657–658).

The present study draws on this rich documentation to investigate the complexities of a pervasive paradigmatic contrast in Jabêm that lies at the intersection of phonology, morphology and syntax: the distinction between realis and irrealis. The terms ‘realis’ and ‘irrealis’ are English renderings of what Dempwolff (1939:12) called *Modus realis* and *Modus imaginativus*, or *Realis* and *Imaginativ* for short. Other linguists working on related languages in the same area have used the terms ‘actual’ and ‘potential’ to refer to the same

---

Joel Bradshaw and Kenneth L. Rehg, eds, *Issues in Austronesian morphology: a festschrift for Byron W. Bender*, 75–85.

Canberra: Pacific Linguistics 2001

© Joel Bradshaw

kind of contrast (see Hooley 1970, Lauck 1980, Muzzey 1979). Dempwolff (1939:12) relates these categories to the familiar European ones by equating the realis with the present, imperfect and perfect, and the irrealis with the future, imperative and subjunctive. Negatives can be either realis or irrealis. Thus, for most purposes, it is sufficient to characterise this contrast as one between nonfuture (realis) and future (irrealis).

A brief explanation with regard to Jabêm phonology is needed before proceeding with the analysis. Jabêm and its closest relative, Bukawa (and possibly also Kela), have developed high–low (H–L) tone contrasts fairly recently in their history (see Bradshaw 1979, Capell 1949, Ross 1993). All syllables of morphemes containing voiced obstruents (*b, d, g*) have predictable L tones. However, L tone is now unpredictable in morphemes where sound changes have obliterated the voicing of ancestral obstruents, as in cases where \**v* was lost and \**z* was devoiced to *s*. L remains the marked member of the tonal set. In the strictest orthographic practice, each syllable with an unpredictable L tone is marked with a grave accent. H tone is unmarked. Although H is prescribed when morphemes contain voiceless obstruents (*p, t, k*), it also occurs in neutral circumstances, where there is (or was) no obstruent to condition the choice of H or L. Neutral consonants include the palatal and velar glides (spelled *j* and *w*), the liquid *l* and the nasals (spelled *m, n, ŋ*). Although the voiceless obstruent *s* now appears to be neutral, since it occurs in both H and L stems, it regularly reflects earlier \**z* when it appears in L-tone stems.

## 2 Historical development

The historical development of the realis/irrealis alternation in Jabêm and its relatives in the Huon Gulf family is fairly well understood. Lynch (1975) outlined a sequence of stages by which a morpheme reconstructible as \**na* ‘irrealis’ first became a verbal prefix (\**na-*), then lost its vowel (\**n-*), then fused with the initial consonant of the verb root (\**n-+\*t- > \*nd-*), eventually giving rise to an oral-grade vs nasal-grade (or lenis vs fortis) alternation between initial consonants in realis vs irrealis verb stems. Ross (1988:360–375) fleshed out Lynch’s schema for a wide range of Western Oceanic languages. One or more of the various stages can be found in Jabêm and many of its closer relatives. Numbami (Bradshaw, fieldnotes), Iwal (Davidson & Davidson 1976:4, 32) and Hote (Ross 1988:371) illustrate the progressive reduction of the irrealis prefix and its incorporation into the verb stem.

	REALIS	IRREALIS	GLOSS
Numbami	<i>ta-tawi</i>	<i>ta-na-tawi</i>	‘we build’
Iwal	<i>ta-tav</i>	<i>ta-n-tav</i>	‘we build’
Hote	<i>a-lav</i>	<i>a-dav</i>	‘we build’

In Jabêm, a handful of earlier vowel-initial roots show remnants of the *-n-* stage in the evolution of the irrealis marker, as in *ta-êc/ta-n-êc* (realis/irrealis) ‘we lie down’, but the regular reflex is now a strikingly innovative suprasegmental that accords well with the other striking phonological innovation of Jabêm—phonemic tone. On L-tone verb stems, the irrealis shows up as prenasalisation, not just of the initial obstruent, but of *all* obstruents in the stem. This leads Ross (1993:139) to claim, “The Yabem irrealis morpheme is prenasalisation, which (like tone) is a feature of the [disyllabic] foot, applying to all available stops and to /s/”.



REALIS	IRREALIS	GLOSS
<i>ta-daguc</i>	<i>ta-ndaŋguc</i>	'we follow'
<i>ta-gadê</i>	<i>ta-ŋgandê</i>	'we cover'
<i>ta-jàba</i>	<i>ta-jàmba</i>	'we work in an awkward position'
<i>ta-màdi</i>	<i>ta-màndi</i>	'we knock'
<i>ta-wàbu</i>	<i>ta-wàmbu</i>	'we entertain to repay'

Ross (1988:156) also presented evidence for reconstructing a realis-marking initial \*g- on singular subject prefixes in Proto Huon Gulf (see also Bradshaw 1979:201–202). This marking is well preserved in Iwal and Jabêm. Note also that Iwal, but not Jabêm, consistently marks the irrealis prefixes with *n-*:<sup>1</sup>

	REALIS	IRREALIS	GLOSS
Iwal	<i>ga-li</i>	<i>na-li</i>	'I see'
	<i>gu-li</i>	<i>nu-li</i>	'you (S) see'
	<i>gi-li</i>	<i>ni-li</i>	'(s)he sees'
Jabêm	<i>ga-lic</i>	<i>ja-lic</i>	'I see'
	<i>gô-lic</i>	<i>ô-lic</i>	'you (S) see'
	<i>gê-lic</i>	<i>ê-lic</i>	'(s)he sees'

Where do all these developments leave realis/irrealis marking in present-day Jabêm? The irrealis is now marked by the addition of nasalisation on L-tone stems with obstruents and on a handful of irregular stems, but it remains unmarked on all other verb stems. At the same time, the realis is marked by an initial velar obstruent on the singular subject prefixes, but remains unmarked on the plurals. Let us now take a closer look at the synchronic status and domain of the realis/irrealis alternation.

### 3 Current status

#### 3.1 Irrealis as a feature of verb stems

The irrealis marker has now been incorporated into Jabêm verb stems and no longer appears as a segmentable prefix. But how regular or reliable a marker is it? Of the 561 verb stems that I counted in appendix 9 of Streicher's dictionary (1982:659–671), less than one-third (29 per cent) show the addition of nasalisation to a separate irrealis form of the stem<sup>2</sup> (see Table 1). All but five of these are L-tone stems containing obstruents that show the addition of homorganic prenasalisation in the irrealis. The five exceptions are all irregular verbs of extremely high frequency that preserve the last traces of the *-n-* prefix on their irrealis stems. All appear to have been vowel-initial at one time, with subsequent accretion of

<sup>1</sup> The following abbreviations are used herein: 1 – first person, 2 – second person, 3 – third person, FUT – future tense, G – genitive, H – high tone, IP – inclusive plural, IR – irrealis, L – low tone, NP – noun phrase, Nu. – Numbami, P – plural, PERF – perfect, POc – Proto Oceanic, R – realis, S – singular, XP – exclusive plural, Ø<sub>i</sub> – zero object pronoun (subscript shows co-reference). R/IR may be shown on the prefix gloss or stem gloss or both. Some of the verbs are glossed with words that are not verbs in English.

<sup>2</sup> Streicher actually lists about 575 verb stems, but I did not count three categories of entries: (i) a few stems listed only for crossreference purposes, such as *-kili*, a variant of *-kêli*; (ii) four *j*-initial stems that contain invariable prenasalised voiced obstruents in both realis and irrealis, such as *-jànda* 'to chase'; and (iii) five *l*-initial stems that show variable prenasalisation in the realis, such as *-lê(n)dêl-lêndê* 'to illuminate (realis/irrealis)'.

palatal glides (Jabêm orthographic *j*) before initial low vowels (compare Jabêm *ja*, Nu. *yawi* ‘fire’ < POc \**api*). Like the other members of their conjugation class (see below), these irregular stems are neutral with respect to tone. They are L-tone when prefixed by realis singular *ga-* ‘1S’, *gô-* ‘2S’, or *gê-* ‘3S’, but H-tone elsewhere.

IRREGULAR STEMS MARKED WITH *-n-* IN THE IRREALIS

REALIS	IRREALIS	GLOSS
<i>ta-eŋ</i>	<i>ta-niŋ</i>	‘we eat’
<i>ta-êc</i>	<i>ta-nêc</i>	‘we lie down’
<i>ta-ja</i>	<i>ta-na</i>	‘we go’
<i>ta-jac</i>	<i>ta-nac</i>	‘we beat, hit’
<i>ta-jam</i>	<i>ta-nam</i>	‘we do, work’

Table 1: Irrealis marking on verb stems

STEM-INITIAL SEGMENTS	IRREALIS ADDS NASAL	IRREALIS = REALIS
<i>b-, d-, g-</i>	115	0
<i>p-, t-, k-</i>	0	133
<i>s- (&lt; *z, *s)</i>	32	46
<i>m-, n-, ŋ-</i>	8	56
<i>w-, j-, l-</i>	8	126
<i>a-, e-, ê-, i-, o-, ô-, u-</i>	2	35
TOTAL	165	396
PERCENT	29%	71%

### 3.2 Realis as a feature of subject prefixes

The absence of any irrealis marking at all on over 70 per cent of verb stems is mitigated by realis marking on the subject prefixes attached to those stems. In fact, the choice of subject prefixes provides the organising principle for the traditional division of Jabêm verbs into five conjugation classes (see Dempwolff 1939, Streicher 1982, Zahn 1940). However, the problem here is that only the singular subject prefixes distinguish realis from irrealis; there is no such distinction in the plural.

Jabêm harmony requirements with regard to obstruent voicing and tone gave rise to two sets of subject prefixes (Bradshaw 1979). The obstruents (except *s*) in the L-tone prefixes are voiced, while those in the H-tone prefixes are voiceless. Thus, alongside its realis singular prefixes with inherited voiced obstruents (L-tone *ga-* ‘1S’, *gô-* ‘2S’, *gê-* ‘3S’), Jabêm has an innovative set of realis singular prefixes with voiceless obstruents (H-tone *ka-* ‘1S’, *kô-* ‘2S’, *kê-* ‘3S’). The innovation went in the other direction for two plural prefixes. Conservative H-tone *ta-* ‘1P’ (cf. Nu. *ta-*, Iwal *ta-*) has an innovative L-tone counterpart *da-*, and conservative H-tone *tê-* ‘3P’ (cf. Nu. *ti-*, Gitua *ti-*) has an innovative L-tone counterpart *dê-*. (The variants *tê-/dê-* ‘3P’ only occur on stems beginning with *t-/d-*, but even there they are being replaced by invariant *sê-* ‘3P’; Streicher 1982:634).

SUBJECT PREFIX SETS		1S	2S	3S	1P	1XP	2P	3P
H-TONE REALIS		<i>ka-</i>	<i>kô-</i>	<i>kê-</i>	<i>ta-</i>	<i>a-</i>	<i>a<sup>3</sup></i>	<i>sê-/tê-</i>
IRREALIS		<i>ja-</i>	<i>ô-</i>	<i>ê-</i>	(plural same as realis)			
L-TONE REALIS		<i>ga-</i>	<i>gô-</i>	<i>gê-</i>	<i>da-</i>	<i>a-</i>	<i>a-</i>	<i>sê-/dê-</i>
IRREALIS		<i>ja-</i>	<i>ô-</i>	<i>ê-</i>	(plural same as realis)			

Table 2: Verb conjugation classes

CONJUGATION	REALIS	IRREALIS	REALIS	IRREALIS	GLOSS
1 H stem	<i>ka-sôm</i>	<i>ja-sôm</i>	<i>ta-sôm</i>	<i>ta-sôm</i>	'I/we speak'
2 L stem	<i>ga-sòm</i>	<i>jà-nsòm</i>	<i>da-sòm</i>	<i>da-nsòm</i>	'I/we search'
3 L/H stem	<i>ga-jà</i>	<i>ja-na</i>	<i>ta-ja</i>	<i>ta-na</i>	'I/we go'
L/H stem	<i>ga-lôb</i>	<i>ja-lôp</i>	<i>ta-lôp</i>	<i>ta-lôp</i>	'I/we fly'
4 HH stem	<i>ka-sapu</i>	<i>ja-sapu</i>	<i>ta-sapu</i>	<i>ta-sapu</i>	'I/we err'
LL stem	<i>ka-sàban</i>	<i>ja-nsàmban</i>	<i>ta-sàban</i>	<i>ta-nsàmban</i>	'I/we wrap'
5 L stem	<i>ka-dôn</i>	<i>ja-ndôn</i>	<i>ta-dôn</i>	<i>ta-ndôn</i>	'I/we teach'

The first conjugation consists of about seventy-seven monosyllabic stems with voiceless obstruents and H tone that take the H-tone set of prefixes. The second conjugation contains about seventy-nine monosyllabic stems with voiced obstruents and L tone that take the L-tone prefixes. The third conjugation is composed of about seventy monosyllabic stems that remain neutral with respect to tone. They take the conservative set of prefixes: L-tone *ga-*, *gô-*, *gê-* in the realis singular, and H-tone *ta-*, *a-*, *sê-* (~ *tê-*) in the realis and irrealis plural. These prefix-conforming stems are thus L-tone in the realis singular, and H-tone elsewhere.<sup>4</sup> Harmony requirements do not apply to the prefixes on disyllabic verb stems, so all 325 disyllabic H-tone and L-tone stems in the fourth conjugation take the now unmarked, H-tone set of prefixes, just as the H-tone stems of the first conjugation do. This conjugation comprises well over half of all Jabêm verb stems. Finally, the tiny fifth conjugation contains twenty-four monosyllabic L-tone stems that violate normal harmony requirements by taking H-tone prefixes. Both Dempwolff (1939:17) and Streicher (1982:632) consider the fifth conjugation to be composed of erstwhile disyllabic stems, in other words, to be composed of former members of the fourth conjugation. However, there may be another factor at work besides syllable reduction. Of the twenty-four (L-tone) stems in the fifth conjugation, nineteen have homophones or near homophones in the second (L-tone) or third (L/H-tone) conjugation. (Compare the following fifth-conjugation [H-L] and second-conjugation [L-L] verbs: *ka-mbu* 'I'll turn [it] on its axis' vs *ga-mbu* 'I'll blame [him]'; *ka-de* 'I groaned' vs *ga-dec* 'I detested'; *ka-nsiŋ* 'I'll hit [it] with a stick' vs *ga-nsiŋ* 'I'll move [it] to and fro'.) The fifth conjugation thus provides both a temporary refuge from homophony and also a beachhead for the analogical spread of the tonologically unmarked (H-tone) set of prefixes to more and more

<sup>3</sup> Although '1XP' and '2P' subjects are not distinguished by the verbal prefixes (both *a-*), they are distinguished by the independent pronouns that occupy the subject NP slot, *aêac* '1P' vs *amàc* '2P'. Conversely, the independent pronouns do not distinguish '1IP' from '1XP' (both *aêac*), but the subject prefixes do, with *ta-* '1IP' vs *a-* '1XP'.

<sup>4</sup> A handful of these neutral stems end in alternating obstruents that harmonise with the voicing of the obstruents in the prefixes, thus *gê-lôb* 'it flew', but *sê-lôp* 'they flew'. Labials (*p*, *b*, *m*) and glottal stop (*c*) are the only consonants that can occur in syllable-final position in Jabêm.

verb stems that now make up the second and third conjugations, the last bastion of the L-tone prefixes, to which barely more than a quarter of verb stems now belong.

**Table 3:** Realis/irrealis marking of verbs in two texts

TEXT	NONE	PREFIX	STEM	BOTH
A	39	17	22	10
B	48	18	19	16
TOTAL	87	35	41	26
PERCENT	46%	18.5%	21.5%	14%

### 3.3 Realis/irrealis as a feature of verbs

The realis/irrealis alternation that pervades Jabêm verbal morphology clearly cannot be assigned to a single morpheme. Explicit irrealis marking shows up on a minority of verb stems (L-tone ones with obstruents) in the form of a suprasegmental feature—prenasalisation—that attaches to every obstruent in the stem. It also shows up as a segmentable stem-initial formant *-n-* on a handful of irregular verb stems. Explicit realis marking, on the other hand, shows up on a minority of subject prefixes (singular ones) in the form of an initial segment *g/k-*. Existing grammatical descriptions of Jabêm have thus wisely adopted a word-and-paradigm approach that assigns the polymorphic realis/irrealis feature to the indivisible inflected verb, not just to the stem or the prefix. Neither Dempwolff (1939) nor Zahn (1940) nor Streicher (1982) speaks of either the realis or the irrealis as a morpheme. Moreover, the (admittedly German-influenced) Jabêm orthographic tradition writes the prefix and stem as one word, without the hyphens that I have scrupulously (and perhaps misleadingly) inserted. The traditional analysis recognises the paradigm, not the morpheme, as paramount in manifesting the distinction between realis and irrealis in Jabêm (see Bender 2000).

Even if we expand the domain of realis/irrealis marking to include the whole word, however, many inflected verbs still fail to show any such marking at all. To test how often this distinction is neutralised in actual discourse, I analysed two narrative texts in the second, revised edition of *Buku Sêsamŋa II*, a reader used in Jabêm schools.<sup>5</sup> Both stories appear to contain a good mixture of action, dialogue, singular and plural participants, and the most common verb stems used in daily discourse. The results were surprising. Nearly half (46 per cent) of all verbal tokens showed no marking whatsoever for realis or irrealis (see Table 3).

Does this indicate that the realis/irrealis distinction is dying out? Not if we examine it in its proper domain—the clause. For, just as the segmental irrealis morpheme *-n-* developed into suprasegmental prenasalisation that shows up in all eligible verb stems, so the lexical distinction between realis and irrealis verbs has now developed into a supralexical feature of clauses that shows up in all eligible verbs. While only 54 per cent of the verbal tokens in the two texts are marked for realis/irrealis, 70 per cent of the clauses contain at least one verb unambiguously inflected for either realis or irrealis (see Table 4).

<sup>5</sup> The first edition compiled and edited by F. Bayer in 1928 was revised and edited in 1955 by M. Lechner and Nêdeclabu Male, the latter a native speaker of Jabêm who worked with H. Zahn on the New Testament translation and school materials during the 1920s and 1930s. *Buku Sê-sam-ŋa* is literally 'book 3P-read-for' (= 'book for reading').

### 3.4 Realis/irrealis as a feature of clauses

One remarkable feature of Jabêm besides phonemic tone is verb serialisation (see Bisang 1986; Bradshaw 1983, 1993; Dempwolff 1939). The defining feature of serial verb constructions (SVC) is the presence of more than one verb per functional clause. Verbs often play roles that adpositions (prepositions and postpositions) play in nonserialising languages. Although not richly endowed with adpositions, Jabêm has quite an adequate supply of clause boundary markers. The ubiquitous conjunctions *ma* ‘and’ and *gebe* ‘say’ introduce coordinate and subordinate clauses, respectively; the propositional negator *atom* ‘not’ marks clause-final position; and relative clauses are usually marked with demonstratives not just at one end, but at both ends. Nevertheless, even the written, carefully edited and bounded clauses of the Jabêm reader that I analysed contained an average of 1.7 inflected verbs each (189 verbs in 115 clauses). Although subject marking can switch within Jabêm serial verb constructions (see Bradshaw 1993), realis/irrealis marking cannot. Two (very slightly truncated) examples from the two texts will illustrate.

(1) REALIS SVC

*Ma awê<sub>i</sub> ... sê-kôc Ø<sub>i</sub> gê-mu gê-mêŋ nê malac kê-tiam.*  
 and woman 3P-take 3S R.3S-back R.3S-come 3S.G village R.3S-again  
 ‘And the woman ... they took her back to her village again.’ (A39)

(2) IRREALIS SVC

*A-kêŋ awê<sub>i</sub> ê-ndêŋ aê ja-kôc Ø<sub>i</sub> ja-mu*  
 2P-put woman IR.3S-IR.to 1S IR.1S-take 3S IR.1S-back  
*ja-na ê-tiam.*  
 IR.1S-IR.go IR.3S-again  
 ‘Give the woman to me and I’ll take her back again.’ (B08)

These two examples contain more verbs than most clauses. In fact, I would concede that example (2) (text B, segment 08) consists of two separate clauses on the grounds that it contains two different agents, ‘2P’ and ‘1S’. Nevertheless, both examples illustrate the consistency of realis/irrealis marking throughout each clause. They also illustrate another important point. Main verbs, such as *-kôc* ‘to take’ and *-kêŋ* ‘to put’ in (1) and (2), are less likely to exhibit realis/irrealis marking than other verbs in the clause. Whereas 70 per cent of all verbal clauses in my sample show explicit marking on at least one verb, only 40 per cent of main verbs are so marked (see Table 4).

**Table 4:** Realis/irrealis marking of clauses in two texts

TEXT	SOME VERB MARKED	NO VERB MARKED	MAIN VERB MARKED	MAIN VERB UNMARKED
A	32	10	13	28
B	38	25	27	32
TOTAL	80	35	40	60
PERCENT	70%	30%	40%	60%

In compiling these statistics, I considered the first verb in the clause to be the main verb unless it was a directional verb that described the motion prelude to the ‘main event’ in the serial verb construction. Examples (3) and (4) are three-verb serial verb constructions where Verb 1 describes the motion prelude, V2 describes the principal action, and V3 situates that

action in a particular location. These examples also provide important clues as to why secondary verbs in Jabêm are more likely to show realis/irrealis marking than main verbs.

## (3) MOTION + ACTION + LOCATION SVC

*Ma sê-ja sê-jac lau Sibôma kê-pi malac langwa.*  
and 3P-R.go 3P-R.hit people S. R.3S-on village old

'And they went and attacked the Siboma people up at the old village.' (A36)

## (4) MOTION + ACTION + LOCATION SVC

*Ja-sê-kwêc malac wakuc gê-êc gwêc.*  
R.go-3P-dig village new R.3S-R.lie sea

'They went and built a new village on the seacoast.' (B56)

Recall that added nasalisation marks the irrealis only on L-tone stems with obstruents and on a handful of irregular stems, while velar-initial prefixes mark the realis only in the case of singular subjects. Secondary verbs are more likely than main verbs to display realis/irrealis marking because (i) a higher proportion of them are either L-tone or irregular, and (ii) a whole class of them are always inflected for singular subjects.

- (i) Two of the most common verbs in Jabêm are irregular L/H-tone *-jal-na* (R/IR) 'to go' and L/H-tone *-mêŋ* 'to come'. They can occur as either main verbs or as directional secondary verbs, and their bare stems can occur as directional prefixes on other verbs, as in (4). The shape of its stem always reveals whether *-jal-na* is realis or irrealis. Two more very frequent secondary verbs are L-tone: *-dêŋl-ndêŋ* (R/IR) 'to reach, suffice' and irregular L/H-tone *-êc/nêc* (R/IR) 'to lie (down)'. The former serves as the equivalent of a dative ('to', as in (2)) or temporal ('at') preposition, while the latter serves as one of the most common ways to add a locative ('at, on') phrase to a clause, as in (4).
- (ii) There is a class of adverbial secondary verbs in Jabêm whose members are always inflected for third person singular subjects. When one of them is the last verb in the clause, it appears to be predicated of the event described by the earlier verb(s) rather than of any particular participant in the event. Examples cited above are the adverbial *-tiam* 'again' in (1) and (2), and the locatives *-pi* 'on' in (3) and *-êc* 'at, on' in (4). In addition, Jabêm has a small class of typologically unusual time-of-day verbs that are always inflected for third person singular subjects, such as *gê-bêc* (R.3S-R.night) 'last night' vs *ê-mbêc* (IR.3S-IR.night) 'tonight' and *ge-leŋ* (R.3S-day) 'this morning, after daybreak', and *e-leŋ* (IR.3S-day) 'tomorrow, on the morrow'. (For further discussion of adverbial serialisation in Jabêm, see Bradshaw 1983:188–189, 1993:152–158 and Dempwolff 1939:82.)

It would appear, then, that realis/irrealis marking is now more pervasive in the verbal syntax of Jabêm clauses than it is in the inflectional morphology of Jabêm verbs. However, its primary locus within the clause remains every bit as elusive as it does in the verb. The morphological feature [ $\pm$ realis] cannot be assigned just to the stem of each verb, nor just to the inflectional prefix. Even if we assign it to the paradigmatic verbal word, the distinction is neutralised on roughly every second verb (46 per cent) in running text. Similarly, even though the syntactic feature [ $\pm$ realis] is overtly marked in over two thirds (70 per cent) of the clauses in our sample texts, it cannot be assigned just to the main verb of each clause, where it is in fact less likely to manifest itself than in one of the secondary verbs.

In the context of morphology at the lexical level, Bender (1998) has observed that ‘the sign gravitates to the word’. Twisting this aphorism to suit present purposes, one can observe that signs of realis/irrealis marking in Jabêm gravitate to the verb, but one can never be sure *how*, *where* or even *if* it will manifest itself in any given verbal clause.

#### 4 Future prospects

Jabêm is not the only language among its close relatives to have begun losing the morphological means to distinguish realis from irrealis on verbs. But the more eroded languages have generally compensated to some degree by relying on separate tense-marking particles before the first verb in the clause.

Hote possesses morphological means similar to those of Jabêm (Muzzey 1979:28–32)—although linguists working on Hote and the Buang languages generally label the distinction ‘actual/potential’ rather than ‘realis/irrealis’. Verb stems beginning with the voiced stops *b* or *d* show the addition of prenasalisation in the irrealis, as in *yaha-dum/ya-ndum* (R/IR) ‘1S-work’. Other verb stems show lenis–fortis alternations in their initial consonants that correlate with realis–irrealis, as in *yaha-valya-mba* (R/IR) ‘1S-make’ or *ha-thik/li-sik* (R/IR) ‘3S-wash’. Separate sets of subject prefixes also correlate with realis/irrealis, as in *yaha-ŋgabom/ya-ŋgabom* (R/IR) ‘1S-cook’ and *ha-ŋgolile-ŋgoli* (R/IR) ‘3S-break’, but the correlation seems to be breaking down, with much inexplicable crossover in actual texts (Muzzey 1979:31, 80–100). As a result, Hote speakers often add separate tense-marking particles as reinforcement, as in *yaha-multe ya-mu* (R/IR) ‘(FUT) 1S-rest’ and *ha-tun/te i-tun* (R/IR) ‘(FUT) 3S-find’. Such tense markers only appear once per clause—before the first verb in serial verb constructions.

The Buang languages farther inland show even greater erosion of subject-agreement and realis/irrealis morphology on verbs. Some Central Buang verbs still show remnants of subject agreement, but only for person and not number. Realis/irrealis is marked by the absence or presence of a *na-* prefix on a few verbs and by root-initial alternations between lenis and fortis consonants in another class of verbs. In many cases, therefore, the appropriate category is “deduced from context, or by the addition of an adverb of time, or by association with another verb which is marked” (Hooley 1995:736). Patep Buang verbs fail to show subject agreement at all except on a tiny handful of vowel-initial stems,<sup>6</sup> and they show no trace of realis/irrealis (or “actual/potential”) morphology (Lauck 1980:9–13). To resolve any resulting ambiguities, Patep speakers must now resort to independent pronouns in subject noun phrase position and a separate irrealis-marking particle, *ob*, in predicate-initial position (1980:22–23). Realis remains unmarked.

##### (5) PATEP REALIS VERB WITH IRREALIS COMPLEMENT (Lauck 1980:74)

*he* Ø *nêb* [*he ob la yêp* ...]  
 3P R want [3P IR go sleep]  
 ‘they decided they would go sleep ...’

Compare an equivalent desiderative structure in Jabêm, where the complement clause is also obligatorily irrealis, but the marking is entirely by means of the morphology of the secondary verb, not by any free-standing morpheme.

<sup>6</sup> The subject-prefixed verbs include *-a* ‘to eat’, *-ê* ‘to see’, *-êp* ‘to sleep’ and *-ib* ‘to die’, and their prefixes only distinguish person, not number: *x-* ‘1S/P’, *w-* ‘2S/P’, *y-* ‘3S/P’ (Lauck 1980:10).



## (6) JABÊM REALIS VERB WITH IRREALIS COMPLEMENT

*se-be* [ *sê-kêng ê-ndêng êsêac* ]  
 3P-R.want [ 3P-give IR.3S-IR.to 3P ]  
 'they wanted to give it to them' (A13)

To compensate for the past and future erosion of the morphological means to distinguish realis from irrealis on verbs, one might expect Jabêm speakers at some point to begin recruiting other means of marking the distinction, as Hote and Patep speakers have done. At this point, the likeliest candidate would seem to be the particle *oc* 'FUT', which according to Streicher (1982:454–455) indicates future tense or probability, as in *kom oc ê-nac* (rain FUT 3S-IR.hit) 'it is going to rain, it may rain', *lau oc sê-mêŋ* (people FUT 3P-come) 'the people will come', and *lau oc sê-mêŋ-gac* (people FUT 3P-come-PERF) 'the people will have come, the people will most likely have arrived'.<sup>7</sup> So far, however, *oc* 'FUT' is not common enough to compensate for those cases (30 per cent) where realis/irrealis remains unmarked anywhere in the clause.

## References

- Bender, Byron W., 1998, The sign gravitates to the word. In Mark Janse with An Verlinden, eds *Productivity and creativity: studies in general and descriptive linguistics in honor of E.M. Uhlenbeck*, 15–25. Berlin: Mouton de Gruyter.
- 2000, Paradigms as rules. In Videia P. DeGuzman and Byron W. Bender, eds *Studies on grammatical analysis*, 14–29. Oceanic Linguistics Special Publications, No. 29. Honolulu: University of Hawai'i Press.
- Bisang, Walter, 1986, Die Verb-Serialisierung im Jabêm. *Lingua* 70:131–162.
- Bradshaw, Joel, 1979, Obstruent harmony and tonogenesis in Jabêm. *Lingua* 49:189–205.
- 1983, Dempwolff's description of verb serialisation in Yabem. In Amram Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol. 4, *Thematic variation*, 177–198. Canberra: Pacific Linguistics.
- 1993, Subject relationships within serial verb constructions in Numbami and Jabêm. *Oceanic Linguistics* 32:133–161.
- Capell, Arthur, 1949, Two tonal languages of New Guinea. *Bulletin of the School of Oriental Studies* 13:184–189.
- Davidson, Ian and Doris Davidson, 1976, Essentials for translation: Iwal language. MS., 50 pp.
- Dempwolff, Otto, 1939, *Grammatik der Jabêm-Sprache auf Neuguinea*. Abhandlungen aus dem Gebiet der Auslandskunde, vol. 50. Hamburg: Reimer.
- Hooley, Bruce A., 1970, Mapos Buang, Territory of New Guinea. PhD dissertation, University of Pennsylvania.

<sup>7</sup> The particle *oc* also appears in the role of 'then' in 'if ... then' (*embe ... oc*) constructions.



- 1995, Central Buang. In Darrell T. Tryon, ed. *Comparative Austronesian dictionary: an introduction to Austronesian studies*, part 1, fascicle 2, 731–740. Berlin: Mouton de Gruyter.
- Lauck, Linda, 1980, Patep grammar sketch. Computer printout, S.I.L., Ukarumpa, Papua New Guinea.
- Lynch, John, 1975, Oral/nasal alternation and the realis/irrealis distinction in Oceanic languages. *Oceanic Linguistics* 14:87–99.
- Muzzey, Margarite, 1979, Hote grammar essentials. Typescript, 100 pp.
- Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of Western Melanesia*. Canberra: Pacific Linguistics.
- 1993, Tonogenesis in the North Huon Gulf chain. In Jerold A. Edmondson and Kenneth J. Gregerson, eds *Tonality in Austronesian languages*, 133–153. Oceanic Linguistics Special Publication No. 24. Honolulu: University of Hawai'i Press.
- Streicher, J.F., 1937, German–Jabêm dictionary. Mimeo, 245 pp.
- 1982, *Jabêm–English dictionary*. Canberra: Pacific Linguistics. (First compiled by Heinrich Zahn in 1917; translated and revised by J.F. Streicher.)
- Zahn, Heinrich, 1917, Wörterbuch der Jabêm-Sprache: Jabêm–Deutsch. Mimeo, 690 pp. Logaweng, New Guinea.
- 1940, *Lehrbuch der Jabêmsprache (Deutsch-Neuguinea)*. Zeitschrift für Eingeborenen-Sprache, supplement 21. Berlin: Reimer.



# *What happened to Erromangan possessive morphology?*

---

TERRY CROWLEY

The possessive systems of the languages of Erromango in Vanuatu exhibit a number of typologically odd features for Oceanic languages in that they do not express the distinction between edible, drinkable (or any other kinds of possession) in alienably possessed nouns, while there are several competing typologically unusual (and semantically equivalent) constructions expressing alienable and inalienable possession. This paper examines the diversity of structural patterns that are found, and argues that the likely explanations involve a combination of imperfect learning on the part of nineteenth century European missionaries, dialect/language leveling between the various Erromangan languages at a time of great demographic change, and possibly influence from an incipient English-lexifier contact language around the same time, and maybe also contact with other languages further to the north.

## **1 Introduction<sup>1</sup>**

Erromango, in southern Vanuatu, has a population of approximately 1,250 people, all of whom speak the language that I will generally refer to simply as Erromangan, though it has also been referred to as Sie /sye/ (Lynch 1983a). This language has now been fairly extensively described (Lynch & Capell 1983, Crowley 1998a).

Erromango was once linguistically more diverse than at present. Massive depopulation in the late nineteenth and early twentieth centuries after initial contact with Europeans resulted in the loss of several languages. There are about half a dozen elderly people on the island who still speak the language known as Ura, which was once more widely spoken in the northern part of the island. While not as well described as Erromangan, a reasonable amount of data has been gathered on the language (Lynch 1983b, Crowley 1998b).

---

<sup>1</sup> Thanks are due to Joel Bradshaw and John Lynch for comments to an earlier version of this paper, as well as those who attended for presentations at the Twelfth New Zealand Linguistic Society Conference at the University of Otago, Dunedin, on 26–28 November 1997, and the Austronesian Circle of Honolulu, on 3 December 1998. Despite these acknowledgements, however, final responsibility for all conclusions in this paper belongs with the present writer.

There was once an additional northern language, known as Utaha, though the last speaker of this language reportedly died in 1954, and virtually no information from the last speakers has survived. The only published information that we have on the language is a short vocabulary and a few grammatical paradigms that were published as Gordon (1889), though I have been able to record a handful of lexical and grammatical items from an Ura-speaker who remembers fragments of Utaha from his boyhood.

The names Sorug /soruj/ (alternating with Sye) and Enyau (alternating with Yocu /yoɣu/) are associated in the nineteenth-century literature with the southern part of the island. It is unclear whether these names represented two distinct (but closely related) languages, or simply represented separate names for dialects of a single language, as the amount of linguistic and sociological information provided in primary sources is so restricted. A more detailed reconstruction of the precontact linguistic geography of Erromango is found in Crowley (1997).

In some respects, the possessive systems of Erromangan languages appear to be fairly typical for Oceanic languages. In particular, the existence of a formal distinction between direct and indirect possessive constructions is exactly as we would expect for languages belonging to this subgroup. In direct possessive constructions, pronominal suffixes are attached to bound nominal roots, which have referents that are typically in an inalienable relationship to the possessor. In indirect possessive constructions, on the other hand, pronominal suffixes are attached instead to postposed possessive constituents. Such constructions typically express a semantically alienable relationship between referents of the possessor and the possessum. Thus, we find a contrast between the following constructions in Erromangan:<sup>2</sup>

<i>noru-ŋ</i>	<i>nimo</i>	<i>horu-ŋ</i>
hand-1SG	house	POSS-1SG
'my hand'	'my house'	

However, there are some aspects of the possessive systems of Erromangan languages which are typologically odd for Oceanic languages in general, or for Southern Vanuatu languages in particular, suggesting that these languages have undergone a significant amount of change in a part of the morphology that is typically fairly stable in Oceanic languages. The main points that are considered noteworthy are the following:

- (i) There is no evidence for separate indirect possessive markers for edible, drinkable and general possession similar to those patterns which are widely distributed in Oceanic languages. Instead, we find evidence for just a single category of indirect possession in all three languages (§2.2, §3.2, §4).
- (ii) Erromangan—but not, it seems, Ura or Utaha—has two separate, but semantically equivalent, paradigms for postposed indirect possessive markers (§2.2).
- (iii) There is some evidence for a separate set of preposed markers of indirect possession based on a separate root (§2.2, §4). Oceanic languages with the VO word order that is characteristic of all Erromangan languages, generally have postposed rather than preposed possessive markers.

<sup>2</sup> Abbreviations used in this paper are NEG – negative, PL – plural, POSS – possessive, PRES – present, SG – singular.

- (iv) Although there is evidence for a set of pronominal suffixes used in the expression of direct possession which are cognate with functionally equivalent forms in other Oceanic languages, there is also evidence in these languages for a competing pattern with directly possessed nouns in which the noun is suffixed by what looks like the third person singular suffix *-n*, to which is attached the free form of the pronoun (§2.1, §3.1). In such constructions, there is little difference between nominal and pronominal possession. This kind of pattern is not otherwise encountered in languages of the Southern Vanuatu, Central Vanuatu or North Vanuatu groupings.
- (v) There is also evidence for an additional construction in which a noun inflected with the third person singular pronominal suffix can be followed by a separate postposed possessive pronoun (§5). This is also not otherwise encountered in descriptions of Vanuatu languages.
- (vi) The languages of southern Vanuatu do not mark number concord between directly suffixed nouns and nominal possessors, except in translated written materials in Erromangan (§5).

All of these features are sufficiently unusual to warrant diachronic discussion. In this paper, I will describe the nature of the changes that have taken place, and I will also make some suggestions as to why these changes might have taken place.

## 2 Erromangan possessive constructions

I will begin by describing the possessive constructions that are found in the modern Erromangan language, as this is the language about which we have the most detailed information.<sup>3</sup>

### 2.1 Directly possessed nouns

With nouns having referents that typically enter into an inalienable possessive relationship with a possessor, the pronominal category of the possessor is expressed by means of suffixes which are attached directly to the noun itself. There is a considerable amount of morphological alternation, and even some irregularity, in the paradigms of directly suffixed nouns, which I will not attempt to describe in this discussion. The regular suffixes found with such nouns are set out in Table 1. Table 2 sets out the regular paradigm for the inflected forms of the root *noru-* 'hand'. When the possessor is expressed as a noun, the possessed noun appears before the possessor noun, with the possessed noun obligatorily carrying third person singular pronominal marking (regardless of the number of the possessor noun), for example:

<i>noru-n itais</i>	<i>noru-n ovn-itais</i>
hand-3SG old man	hand-3SG PL-old man
'the old man's hand'	'the old men's hands'

<sup>3</sup> For the sake of coherence, some details relating to possessive constructions are not presented here. A more complete account is presented in Crowley (1998a).

**Table 1:** Possessive pronominal suffixes in Erromangan

	Singular		Plural
First person	-ŋ	Inclusive	-(n)t
		Exclusive	-mam
Second person	-m		-mi
Third person	-n		-nr

**Table 2:** Possessive paradigm for *noru-* 'hand' in Erromangan

	Singular		Plural
First person	<i>noruŋ</i>	Inclusive	<i>noru(n)t</i>
		Exclusive	<i>norumam</i>
Second person	<i>norum</i>		<i>norumi</i>
Third person	<i>norun</i>		<i>norunr</i>

There is a competing pattern of direct possession with pronominal possessors in the first and second persons, in which the suffixes take the forms set out in Table 3. We therefore find alternation with second person singular possessors between *noru-m* and *noru-nkik*. While these both mean 'your hand', the first alternative (with the suffix shown in Table 1) is the one most frequently encountered, while the second (with the suffix shown in Table 3) is associated with the expression of contrast; thus:

*noru-ŋ*    *yam-nohoru*    *kou*    *noru-nkik*    *yotum-nohoru*  
 hand-1SG    3SG:PRES-sore    but    hand-2SG    3SG:NEG:PRES-sore  
 'My hand is sore but yours (in contrast) is not.'

**Table 3:** Alternative possessive suffixes in Erromangan

	Singular		Plural
First person	-nyau	Inclusive	-nkoh
		Exclusive	-nkam
Second person	-nkik		-nkimi
Third person	-n		-nr

The suffixes in this contrasting paradigm are, in fact, morphologically complex. Compare Table 3 with Table 4, which lists the independent pronouns in Erromangan. The first and second person suffixes in Table 3 involve the third person singular form of the noun, to which the corresponding first and second person pronouns have been phonologically bound. There is, therefore, a close structural parallel between this second possessive construction and the pattern described above for nominal possessors. The only difference is that nominal possessors are stressed as separate phonological words, whereas pronominal possessors are treated as part of the same word as the preceding possessed noun for stress assignment purposes; thus:

*noru-n*    *itais*    *noru-n-yau*  
 hand-3SG    old man    hand-3SG-1SG  
 'the old man's hand'    'my hand'

**Table 4:** Independent pronouns in Erromangan

	Singular		Plural
First person	<i>yau</i>	Inclusive	<i>koh</i>
		Exclusive	<i>kam</i>
Second person	<i>kik</i>		<i>kimi</i>
Third person	<i>iyi</i>		<i>iror</i>

## 2.2 Indirectly possessed nouns

When a free-form noun refers to an entity that is possessed alienably, there are two separate possessive pronouns which can be postposed to the noun. The first of these possessive markers has the paradigm set out in Table 5 for the full range of pronominal possessors. The first and second person forms of this paradigm are based upon a root that is indeterminate between underlying *hore-* and *horo-*. This root accepts pronominal suffixes of the same shape that we find with directly suffixed nouns, with a number of minor exceptions. It should be noted that the third person forms here are based on a completely different root, namely *ihe-*, though the suffixes are exactly as we would expect: *-n* '3SG' and *-nr* '3PL'.

**Table 5:** Possessive postmodifiers (Set I) in Erromangan

	Singular		Plural
First person	<i>horuŋ</i>	Inclusive	<i>horet</i>
		Exclusive	<i>hormam</i>
Second person	<i>horom</i>		<i>hormi</i>
Third person	<i>ihen</i>		<i>ihenr</i>

The paradigm for the second series of these possessive postmodifiers is as set out in Table 6. There is no semantic contrast between the use of *e-* and *he-* with these possessive constituents, though the vowel-initial forms are more frequently encountered. The final elements of most of these possessive pronouns are identical in shape to the free pronouns. The only exceptions are the third person singular form, which ends in *-i* rather than *-iyi*, and the second person plural form, which ends in *-kmi* rather than *-kimi*. The roots in this paradigm alternate between *(h)enoŋ-* when the suffixed element begins with *k*, and *(h)en-* otherwise.

**Table 6:** Possessive postmodifiers (Set II) in Erromangan

	Singular		Plural
First person	<i>(h)enyau</i>	Inclusive	<i>(h)enoŋkoh</i>
		Exclusive	<i>(h)enoŋkam</i>
Second person	<i>(h)enoŋkik</i>		<i>(h)enoŋkmi</i>
Third person	<i>(h)eni</i>		<i>(h)eniror</i>

These possessive pronoun paradigms are used completely interchangeably with all nouns that participate in the indirect possessive construction. None of the semantic contrasts that are so frequently encoded in Vanuatu languages by different forms of the possessive constituent, such as edible possession and drinkable possession, is expressed grammatically in this language; thus:

*nimo enyau*  
house POSS:1SG  
'my house'

*nimo henyau*  
house POSS:1SG  
'my house'

*nimo horu-ŋ*  
house POSS-1SG  
'my house'

There is one additional form in Erromangan having a possessive function and that is *naŋku* 'my'. This differs structurally from the other first person singular possessive pronouns *horuŋ* and *(h)enyau* in that it appears before the possessed noun rather than after it. The construction below is therefore semantically equivalent to the three forms just presented:

*naŋku nimo*  
POSS:1SG house  
'my house'

The proposed possessive form is further unusual in that it is not paradigmatically related to forms expressing other pronominal distinctions in the possessor. Thus, while one might have expected, on historical grounds at least, to find forms such as \**namu* 'your', this is not recognised at all by speakers of the language.

The constructions described above all involve pronominal rather than nominal possessors. When an indirectly possessed noun is associated with a nominal possessor, the possessed noun phrase is followed by the possessor noun phrase, and the two are linked by either *(h)en* or *ihen* in free variation;<sup>4</sup> thus:

<i>nimo (h)en natmonuy</i>	<i>nimo ihen ov-atmonuy</i>
house POSS chief	house POSS PL-chief
'the chief's house'	'the chiefs' houses'

It should be noted that while *ihen* can function as both a third person singular possessive pronoun and a marker of nominal possession, the third person singular possessive pronoun *(h)eni* is slightly different in shape to the marker of nominal possession, which is *(h)en*.

### 2.3 Possessive constructions in written literature

The printed literature of Erromangan is not extensive in scope, and is entirely religious in content. The lengthiest printed document is the *New Testament*, which was translated in the late nineteenth and early twentieth centuries by a series of English-speaking Presbyterian missionaries. There is also a printed catechism and hymnal, which was originally produced in 1867, and which has been reprinted—largely without change—many times since then.

The vast majority of printed materials in Erromangan have been written by foreigners. There are many respects in which structures in these translated materials show evidence of learner errors. Interestingly, many of these features have been replicated in written Erromangan texts, particularly hymns, written in recent years by native speakers of Erromangan, who have evidently come to regard these earlier translations as exemplars of high ecclesiastical style to be emulated in written religious texts.

My corpus of spoken Erromangan contains many instances of the indirect possessive constructions presented in §2.2, as well as the direct possessive construction described in §2.1, involving the pronominal suffixes set out in Table 1. While a contrast is made in my

<sup>4</sup> Although there is a productive synchronic rule of vowel-initial deletion in Erromangan, the alternation between *hen* and *ihen* does not fit the structural description for the application of this rule. This alternation is, therefore, not part of any general synchronic process.



spoken corpus between directly suffixed possessive constructions such as *noru-ŋ* ‘my hand’ and *noru-nyau* ‘my hand’, this is a distinction that is not made at all in printed materials in the language. In these printed materials, only the latter construction is found. This means that the semantic contrast in the spoken style that is associated with the difference in these constructions is lost in written styles.

### 3 Ura

Ura also makes a distinction between directly and indirectly possessed nouns, though the constructions that we find are in some respects significantly different to what we find in Erromangan.<sup>5</sup>

#### 3.1 Directly possessed nouns

The vast majority of directly suffixed nouns in Ura end in the segment *-n*, which is not separable from the rest of the noun. With such nouns, a nominal possessor directly follows the possessed noun, with no intervening constituent of any kind, as in:

*nobun nalinowe*  
 head dog  
 ‘dog’s head’

When such nouns are associated with the expression of pronominal possessors, the independent pronouns are phonologically bound to the noun itself. A third person singular pronominal possessor is irregularly expressed by means of zero. The resulting paradigm for the directly suffixed noun *nobun* ‘head’ is set out in Table 7.

Table 7: Possessive paradigm for *nobun* ‘head’ in Ura

	Singular		Plural
First person	<i>nobun-yau</i>	Inclusive	<i>nobun-gis</i>
		Exclusive	<i>nobun-gim</i>
Second person	<i>nobun-ga</i>		<i>nobun-ŋimi</i>
Third person	<i>nobun-Ø</i>		<i>nobun-leil</i>

Although the set of directly suffixed nouns in Ura is quite large, there is a subset of about half a dozen such forms where there is synchronic evidence for the separability of the final *-n* of the noun. The noun *maran* ‘maternal uncle’ is a noun of this subset, and with this noun the third person singular and the plural forms behave exactly as described above. The first and second person singular forms, however, are based on a root of the shape *mar-*, to which the suffixes *-k* ‘1SG’ and *-m* ‘2SG’ are added. We therefore find paradigms such as that set out in Table 8.

<sup>5</sup> Once again, some minor patterns have been ignored in the following summary of Ura possessive constructions. Note that *b*, *d* and *g* in Ura represent voiced prenasalised stops, corresponding to homorganic nasal-voiceless stop clusters in Sye.

**Table 8:** Possessive paradigm for *maran* ‘maternal uncle’ in Ura

	Singular		Plural
First person	<i>mara-k</i>	Inclusive	<i>maran-gis</i>
		Exclusive	<i>maran-gim</i>
Second person	<i>mara-m</i>		<i>maran-ŋimi</i>
Third person	<i>maran-∅</i> <sup>6</sup>		<i>maran-leil</i>

### 3.2 Indirectly possessed nouns

Pronominal possessors with indirectly possessed nouns in Ura are expressed by the postposed forms set out in Table 9; thus:

*suŋai ar(y)au*  
house POSS:1SG  
‘my house’

**Table 9:** Possessive postmodifiers in Ura

	Singular		Plural
First person	<i>ar(y)au</i>	Inclusive	<i>arkis</i>
		Exclusive	<i>arkim</i>
Second person	<i>arka</i>		<i>arŋimi</i>
Third person	<i>ar(y)i</i>		<i>ahleil</i>

As with the second set of possessive postmodifiers described in §2.2 for Erromangan, these forms also bear a formal similarity to the independent pronouns, which have the shapes set out in Table 10. Apart from the optional loss of *y* in *ar(y)au* ‘my’ and the reduction of *iyi* to either *yi* or *i* in *ar(y)i* ‘his/her/its’, the other alternations here appear to be what we would expect with the addition of these pronouns to an initial element of the shape *ar-* according to the general morphophonemic rules of the language.

**Table 10:** Independent pronouns in Ura

	Singular		Plural
First person	<i>yau</i>	Inclusive	<i>gis</i>
		Exclusive	<i>gim</i>
Second person	<i>ga</i>		<i>ŋimi</i>
Third person	<i>iyi</i>		<i>leil</i>

Nominal possessors with indirectly suffixed nouns are expressed with the possessed noun appearing before the possessor noun, and the two are linked by means of the possessive marker *ar*, as in:

*suŋai ar yarumne*  
house POSS chief  
‘the chief’s house’

<sup>6</sup> This form is obviously ambiguous between the analysis presented in this table, and the root *mara-*, to which a third person singular suffix of the shape *-n* has been added.

It should be noted that in Ura, there is only a single paradigm of postposed possessive pronouns, in contrast to the two synonymous paradigms that we find in Erromangan. There is also no form in Ura corresponding to the preposed first person singular possessive pronoun *nanku* in Erromangan.

#### 4 Utaha

The information on the possessive system of Utaha is restricted to the brief observations in Gordon (1889:79), as well as a handful of forms recorded by myself. Like Erromangan and Ura, this language also appears to have had postposed possessive pronouns used in association with indirectly possessed nouns. The precise forms of the full paradigm are somewhat uncertain. Gordon (1889:79) recorded the form set out in Table 11 in his own orthography. Given Gordon’s orthographic conventions, these forms would appear to be based on an initial element of the basic shape *eti-* or *ete-*, with a remaining element that is identical in shape to the free-form pronouns, for which Gordon (1889:79) provides the paradigm in Table 12 (presented once again in his own orthography).

**Table 11:** Possessive postmodifiers in Utaha

	Singular		Plural
First person	<i>etiyo</i>	Inclusive	<i>etiḡgis</i> <sup>7</sup>
Second person	<i>eteko</i>	Exclusive	<i>etekum</i>
Third person	<i>et iyi</i>		<i>etekimi</i>
			<i>eteyoril</i>

**Table 12:** Independent pronouns in Utaha

	Singular		Plural
First person	<i>yo</i>	Inclusive	<i>ḡgis</i>
Second person	<i>kō</i>	Exclusive	<i>kum</i>
Third person	<i>iyi</i>		<i>kimi</i>
			<i>yoril</i>

I am somewhat reluctant to reconstruct the full paradigms on the basis of Gordon’s published forms because of discrepancies between his recordings of the first and second person singular pronouns, and my own recording of *etyou* and *etko* respectively. My recordings suggest that Gordon’s paradigm may involve illicit epenthetic vowels, and this is a recurring feature of his materials in Erromangan languages. It is therefore possible that the paradigm in Utaha was based on the initial element *et-*, as suggested in Table 13.

<sup>7</sup> In Gordon’s orthography, *ḡg* represented /g/.

**Table 13:** Possible phonemic forms of possessive postmodifiers in Utaha

	Singular		Plural
First person	<i>etyou</i>	Inclusive	<i>etgis</i> <sup>8</sup>
		Exclusive	<i>etkum</i>
Second person	<i>etko</i>		<i>etkimi</i>
Third person	<i>etiyi</i>		<i>etyoril</i>

Gordon (1889:84) also provides a hint of a possible additional indirect possessive paradigm in Utaha in the phrase which he represents orthographically as *namu nelō* 'your kingdom'. The form that is spelt *nelō* is presumably cognate with Erromangan *lou* 'kingdom', to which a second person singular possessive constituent of the shape *namu* has been preposed.

In the absence of any further data, this form might have been treated as some kind of error given the lack of any further exponents of such a paradigm. However, I indicated in §2.2 that *naŋku* 'my' appears as an occasional preposed possessive marker in Erromangan, in addition to other indirect possessive markers which are exclusively postposed. The occurrence of these forms in Erromangan and Utaha together suggest a possible paradigm based on an initial element *na-*, with following *-ŋku* '1SG' and *-mu* '2SG'.

Gordon (1889:79) also provides some hints for the existence of a directly suffixed paradigm in Utaha. In particular, the forms in Table 14 are possible direct possessive suffixes in this language which can be extracted from his data.

**Table 14:** Possible pronominal possessive suffixes in Utaha

	Singular		Plural
First person	<i>-ɣ/-∅</i>	Inclusive	?
		Exclusive	<i>-kum</i>
Second person	<i>-m</i>		<i>-mim</i>
Third person	<i>-n</i>		<i>-ra</i>

Most of these forms are problematic, each in different ways. The only example that Gordon (1889:79) provides of a possible first person singular directly suffixed form is spelt *niseko* 'for me'.<sup>9</sup> Given his orthographic frailties, the suffix here could be interpreted as either zero or *-ɣ*. The form *nisekum* was glossed as both '2SG' and '1PL:EXCL'. This seems an unlikely collapsing of categories, and we would expect that the first person plural exclusive suffix was more likely to have been *-kum* rather than *-m*, given the shape of the free-form pronoun.

Gordon provides no examples which point to the separability of final *-n* on any directly suffixed noun in Utaha, though the form that is written as *timen* 'father' (which I have recorded as /timin/) is cognate with widely distributed Oceanic forms with the same meaning, in which the final *-n* is clearly cognate with the widespread third person singular possessive suffix *-n*. The spelling *nisekira* 'for them' suggests the possibility of morphophonemic alternations in the shape of the preceding root vowels in Utaha associated with different suffixes similar to what we find in Erromangan, though with such a limited amount of information the details cannot be recovered.

<sup>8</sup> It is not known, however, if this represents a plain or prenasalised stop in Utaha.

<sup>9</sup> Although this is a preposition rather than a noun, the cognate form in Erromangan is the benefactive preposition *nis yo-*, which accepts possessive suffixes in the same way as a directly suffixed noun.

## 5 Other unusual patterns

The corpus of published materials in Erromangan languages provides evidence of another pattern of possessive marking beyond those presented in §2, §3 and §4. In my corpus of spoken Erromangan, the noun *ni-* 'name' is recorded as behaving like a perfectly ordinary directly suffixed noun. We therefore find constructions such as the following:

<i>ni-m</i>	<i>ni-n</i>
name-2SG	name-3SG
'your name'	'his/her name'

However, there are repeated examples in nineteenth-century printed materials of the shape *nin soro-m* 'your name'.<sup>10</sup> In this construction, the noun carrying the third person singular suffix is followed by the inflected form of one of the possessive constituents that is associated with the expression of indirect possession.

This construction therefore involves the expression of two possessors, and translates literally as 'your his/her name'. Speakers of Erromangan today comment on this construction in written materials as being 'odd', yet it is widely encountered in nineteenth-century printed materials, and it also appears in more recently composed hymns produced by native speakers of Erromangan.

A similar kind of construction is also noted for Utaha (for which we obviously have no spoken corpus). Gordon (1889:84) presents an example in his own orthography as *nin eteko* 'your name'. His *eteko* corresponds to what I have recorded as *etko*. Assuming that his *nin* 'name' is cognate with Erromangan *ni-* 'name', it is likely that this example also involves the double expression of possessor.

The written translated corpus also contains evidence of another direct possessive construction that is never attested in my spoken corpus, in which there is concord for number between the possessed noun and a following nominal possessor. In my spoken corpus, there is only evidence for third person singular marking on such nouns. Thus spoken

<i>noru-n</i>	<i>ovn-itais</i>
hand-3SG	PL-old man
'the old men's hands'	

corresponds systematically to written

<i>noru-nr</i>	<i>ovn-itais</i>
hand-3PL	PL-old man
'the old men's hands'	

Erromangans regard this latter construction as an error on the part of the European translators, and do not emulate it in their recently composed hymns. Given the lack of such number concord in possessive constructions in the other languages of Southern Vanuatu, this interpretation would seem to be correct. Although this construction does not represent a genuine pattern in Erromangan, I have included reference to this error as the question of missionary errors and their influence on the development of possessive constructions in the modern language is revisited in the following section.

<sup>10</sup> Note that modern *h* regularly derives from nineteenth-century *s*.

## 6 Historical discussion

In this section, I will offer an account of the various kinds of changes that have affected possessive systems in the languages of Erromango, in the light of the typologically unusual features that are found in these languages as summarised in §1.

### 6.1 Developments affecting direct possessive constructions

Of the two main patterns of direct pronominal possession, it is clear that the paradigm represented in modern Erromangan by the forms set out in Table 1 is inherited from Proto Oceanic, while the pattern in Table 3 represents a later innovation. This innovative pattern closely parallels that used in the expression of nominal possession in the languages of Erromango.

Although the early missionaries were aware that there was a structural difference between the expression of alienable and inalienable possession, the semantic contrast between the two patterns of direct suffixation in Erromangan is subtle enough that it would be possible to imagine some linguistically naive missionaries in the nineteenth century not being able to work the details out. Given the comment in §5 that errors were clearly made in the expression of some possessive categories in Erromangan, it is plausible to argue that early translators, when confronted with competing forms seemingly expressing the same meaning, chose the construction that offered the greater amount of morphological transparency as the preferred pattern in their translations. This would account for the fact that in printed materials, only the typologically more unusual—and semantically marked—pattern of direct possession is found.

It is also worth considering the possibility that the early translators might have been attempting to deliberately find structures that would be recognised by speakers of all Erromangan languages. Effectively, then, they would have been moving towards an artificially created written koine. We have documentary evidence that the early missionaries were familiar with Ura as well as modern Erromangan (Robertson 1902:135–36). Ura has all but abandoned the original pattern of direct pronominal possession in favour of the more transparent pattern whereby nominal and pronominal possession are expressed similarly. This may have provided some additional motivation to the early translators to operate exclusively with the innovative pattern in translated materials in Erromangan in order to allow for maximum access to these materials by speakers of Ura at the time.

There is another area in which deliberate mission grammatical adventurism may have played a role in determining the kinds of possessive constructions which appear in translated texts, that is, the double expression of indirect possession noted in Erromangan and Utaha. Interestingly, the only examples of this structure in Erromangan involve the noun *ni-* ‘name’ (along with the same root in Utaha). In Erromangan, this form is homophonous, meaning both ‘name’ and ‘breast’.<sup>11</sup> It is clear from Gordon’s (1889) published notes that he was aware of the morphological distinction between directly suffixed possession and indirect possession, so the construction that he used with this noun is unlikely to have been simply an accidental mistake.

It is tempting to speculate that Gordon was aware of the homophony here, and that he was anxious that a grammatically correct translation of ‘Our Father, who is in heaven, hallowed be

<sup>11</sup> In Proto Southern Vanuatu (Lynch, pers. comm.) there were originally two quite different forms that have coincidentally merged with the same phonemic shape in modern Erromangan.

your name', should not be misinterpreted by recently converted Christians as 'Our Father, who is in heaven, hallowed be your breasts'. He may have deliberately mangled the morphology of the language away from expected *ni-m* 'your name/breasts' to ungrammatical *ni-n soro-m* 'your his/her name/breasts' in the hope that in doing this he might deflect people away from such a profane interpretation.<sup>12</sup>

## 6.2 Developments affecting indirect possessive constructions

Of changes affecting indirect possessive constructions, the loss of separate markers for edible and drinkable possession seems to be a perfectly unremarkable sort of development. Although the other languages in the Southern Vanuatu subgroup, to which Erromangan belongs, have retained separate markers for these categories, there are other Oceanic languages—including those of Efate immediately to the north—which have lost the distinction and use a single set of markers for all categories of indirect possession. This merger of possessive categories would appear to have taken place relatively recently, presumably at the Proto Erromangan stage.

We also need to account for the fact that modern Erromangan—but not, it seems, Ura or Utaha—is the only Oceanic language that is known to have two phonologically unrelated, and structurally quite different, possessive postmodifier paradigms which are, however, semantically equivalent. The forms in Tables 6, 7 and 13 all involve the same basic pattern and are based on the possibly cognate initial elements *en-* in Erromangan, *ar-* in Ura and *et-* in Utaha, with the forms in Table 5 (based on *horo-/hore-*) in Erromangan representing a pattern that appears not to have cognates in the other languages.

A number of hypotheses regarding the source of these competing forms can be proposed. One possibility to consider is that one of the two sets of forms comes from an original general possessive marker, while the other comes from an earlier edible or drinkable possessive marker. The original opposition could then have been merged, with the two sets of markers simply coming to be used interchangeably for all types of alienable possession. While this solution is at first glance plausible, it lacks further appeal because there are no convincing cognates for either of these sets of forms in the edible or drinkable possessive paradigms elsewhere in the languages of the Southern Vanuatu subgroup, or even further afield in the Oceanic subgroup.

Another possibility is that Gordon had simply mislearned Erromangan and that he effectively created a new possessive construction in his written translations, which then became the basis for a general innovation in the language. In Gordon (1889), the forms of the pronouns which appeared in the earliest version of the Lord's Prayer—which he produced in the early 1870s—were those from the morphologically more transparent (*h*)*en(oŋ)-* set. By the time that Gordon's successor, Robertson, presented his revised version of the Lord's Prayer in the same language in 1902, the (*h*)*en(oŋ)-* possessive forms had been systematically replaced by the precursors to the *hore-/horo-* forms (which at that time had the shape *sore-/soro-*). In saying of Gordon that "his knowledge of Erromangan was perfect, and his translations almost without a mistake", Robertson (1902:135) is implicitly claiming that

<sup>12</sup> My suggestion on this point is unproven, and in all likelihood, probably also unprovable. It would also be possible to argue that deliberate morphological intervention represents an unlikely explanation here, as producing an apparently ungrammatical structure could be seen as drawing more attention to the possibility of ambiguity, rather than deflecting people away from an unacceptably profane interpretation.



Gordon made **some** mistakes, and he presumably saw to it that these were subsequently corrected. In this particular text, the substitution of one possessive paradigm for another is the most obvious change of all.

However, it is difficult to take this argument too far as well. It is clear that Gordon did make structural mistakes in his translations, and that some of these have become established as part of a new ecclesiastical register in the language which is actively emulated by Erromangans today when writing hymns of their own. However, there is no evidence that any of these mistakes have been incorporated into the ordinary spoken language. It seems unlikely, therefore, that the appearance of the *(h)en(oŋ)*- possessive postmodifiers in spoken Erromangan could be due to imperfect learning on the part of the early missionaries.

In any case, although the *(h)en(oŋ)*- possessive forms are indeed morphologically more transparent than the *hore-/horo-* forms, they do not behave morphologically exactly as we find with nominal possessors, as noted in §2.2. If Gordon were responsible for the creation of a new paradigm, it would be impossible to account for the origin of the sequence *-oŋ-* in forms such as *enoŋkik* 'your (sg)'.<sup>13</sup> If Gordon were making a mistake, he would surely have regularised this simply as *enkik*, based on the pattern of forms such as *enyau* 'my'.

Additionally, it is only the *(h)en(oŋ)*- possessive forms which appear to have cognates in both Ura and Utaha. Although these forms clearly represent recent innovations, they are clearly widely enough distributed in the languages of Erromango to derive from a pre-contact innovation, which possibly goes back as far as Proto Erromangan.

A similar scenario, which is rather more favourable to the memory of Gordon, is one in which Erromangans may have exposed him to only the *(h)en(oŋ)*- possessive forms. Because of their transparency, they may have felt these to be more 'learnable' than the *sore-/soro-* forms. The experience of the missionary Lawes when he was first exposed to the Motu language of Port Moresby in Papua New Guinea provides something of a parallel here. Lawes put considerable effort into his biblical translation work, though it was some time before he learned from his own children who were growing up speaking Motu with the local children that the adults had been teaching him a structurally simplified foreigner-talk variety of the language (Dutton 1985:44–46).

Another possible explanation for the unusual transparency of the *(h)en(oŋ)*- paradigm—and similar paradigms in Ura and Utaha—is that these forms may have arisen as recent Erromangan calques on Bislama constructions. Bislama is an English-lexifier pidgin in which nominal and pronominal possessive constructions are expressed in exactly the same way, as illustrated by the following:

<i>dog blong olfala</i>	<i>dog blong mi</i>
'the old man's dog'	'my dog'

Erromango underwent massive depopulation in the nineteenth century, which resulted in major demographic realignments. People moved from inland villages to coastal villages

<sup>13</sup> Lynch (pers. comm.) suggests a historical explanation for the origin of the sequence *-oŋ-* which would be consistent with forms such as *enoŋkik* having considerable historical antiquity. He argues that the root of the possessive marker may have been *enə* and that the pronoun may have originally begun with the cluster *ŋk-*. There is a regular synchronic process whereby underlying schwa is realised preconsonantly as *o*, and there would have to have been sporadic loss of the initial nasal with the independent pronoun, thus *enə-ŋkik* > *enoŋkik*. Although such an analysis may be historically valid, and the rule affecting underlying schwa is synchronically valid, there is no independent synchronic motivation for the rule of initial nasal-deletion with independent pronouns.



under missionary pressure, and villages merged as they became unviable through loss of population. This period of major demographic upheaval coincided with some major linguistic upheavals. Not only did Utaha and Ura begin along the road to language death during this time, but this was also the period in which Bislama was becoming established on the island, in association with the commencement of the operations of sandalwood cutters and sea slug processors and the concomitant importation of labourers from other islands in southern Melanesia.

An argument could be made that there may have been some disruption to normal language transmission which may have allowed a calque such as this to become incorporated, as we might expect from Mühlhäusler's (1996) contention that Oceanic languages are currently losing their structural integrity in the face of contact with colonial languages.

The weakness in this argument with regard to the Erromangan languages is essentially the same as the weakness in the argument presented above regarding the possibility that Gordon has simply mislearned the original system. If this construction were based on a Bislama calque, we need to explain why we now find forms such as *enɔŋkik* 'your' in Erromangan, rather than a more transparent form such as *enkik* 'your'. In any case, if there were substantial disruption to normal language transmission, we would expect to find other evidence of calquing or structural simplification in modern Erromangan, which is simply not there, contrary to the grammatically naive generalisations in Mühlhäusler (1996).

Another explanation for the presence of two sets of possessive markers in Erromangan is that the two different forms may derive from markers in two different languages (or two dialects of a single language). Such an interpretation implies that some kind of linguistic amalgamation took place resulting in the development of a language that is structurally mixed, at least with respect to the expression of indirect possession.

There is documentary evidence for this kind of interpretation, and this is the solution proposed by Lynch (1983a:8). Gordon (1889) says that the difference between the speech forms known in the mid-nineteenth century as Enyau and Sorug was 'slight', though we do not know the precise extent of these differences. Lynch (1983a) argues that these speech forms were probably distinct, though closely related, languages. My own reading of the primary sources is that the evidence is indeterminate as to whether these two speech forms were separate languages or just different dialects of a single language (Crowley 1997). However, whether Enyau and Sorug were simply dialects of a single language or distinct but closely related languages is not especially relevant to the discussion which follows.

The names Enyau and Sorug that Gordon used to refer to these speech forms—namely, the respective possessive postmodifiers meaning 'my'—capture what was clearly felt to be a major point of difference between the two. The precursor of the modern series of possessive markers based on the root *hore-/horo-* could therefore be said to derive from the earlier Sorug forms, while the forms based on the root *(h)en(oy)-* could be said to be derived from the earlier Enyau forms. Even in Gordon's time, there is some evidence that there was some alternation between the two paradigms, as the purportedly northern form *eni* 'his/her/its' appears in Gordon's examples alongside the supposedly southern form *isen*.

A possible weakness in this argument is that in situations where there is sufficient contact between dialects or closely related languages for there to be interchange of basic structural features, I would expect a certain amount of structural simplification and levelling to take place, such as we typically find in koineised languages. In Crowley (1995), I describe a number of aspects of the verbal morphology of modern Erromangan in which there is ample

opportunity for developments of this type to have taken place, but there is no evidence that there have been any such changes.

Each of the last three suggestions—missionary intervention via imperfect learning, calquing on incipient Bislama, and internal dialect/language levelling—has some appeal, though each also has its limitations, as I have just indicated. These scenarios are, in fact, not mutually incompatible. Perhaps the real answer is that there may have been an element of imperfect learning on Gordon's part, which conspired with a situation that opened up the possibility of Bislama calquing, as well as the promotion of koineisation, to result in the situation that we find today with regard to the expression of indirect possession in Erromangan.

The fact that Robertson replaced all of Gordon's innovative (*h*)*en(oy)*- possessive forms with the conservative *sore-/soro-* forms could be explained by the opening up of the whole island to Christianity. This would have brought about greater contact with the demographically dominant south, thereby giving a chance for the southern *sore-/soro-* forms to spread the north, according to the koineisation scenario.

There is one further point that needs explanation, and that is the presence of preposed *naŋku* 'my' in modern Erromangan and *namu* 'your' in Utaha. With so little data on Utaha, we are very much restricted in what can be said about this form. One unanswerable question that arises is whether or not there was a complete additional paradigm in Utaha, of which Gordon has provided us with only a single exponent. If there was such a paradigm, we can have no idea about how forms such as preposed *namu* 'your' were related to postposed *etko* 'your'. They may have expressed different categories of indirect possession or they may have been semantically equivalent variants, just as we find with the two major paradigms of indirect possessive markers in modern Erromangan.

Of these forms, *etko* and *namu* meaning 'your', the latter looks more like what we would expect from a language belonging to the Northern and Central Vanuatu subgroup, and the same is true of *naŋku* 'my' in modern Erromangan as against (*h*)*enyau* or *horuŋ*. An earlier form such as \**naŋku* would have ended up in modern Erromangan as *naŋ* if it had been directly inherited. Utaha was spoken in the northern part of Erromango, which opens up the possibility this paradigm may have been introduced as a result of contact with languages from Efate and further north.

Another possibility is that the possessive forms based on the root *na-* represent the original Utaha forms and that the forms based on the root *et-* replaced these under the influence of Ura. Lynch (1995) points to a number of features in Erromangan languages which appear to be transitional between the languages of the Central and Northern Vanuatu subgroup and those of the Southern Vanuatu subgroup. The earlier presence of a paradigm such as this in Utaha may have been another feature of this type, which was subsequently almost completely replaced by a more Erromangan-looking pattern.

There is one fact about the use of the preposed form *naŋku* 'my' in modern Erromangan that still puzzles me. Not only is this form much less frequent than the postposed possessive forms *horuŋ* and (*h*)*enyau*, but it is especially frequently encountered in association with the two borrowed words *poi* 'boy' and *kel* 'girl'. Older people addressing younger people—and even people of similar ages addressing each other in a familiar way—frequently do so using the phrases *naŋku poi* 'my boy'<sup>14</sup> and *naŋku kel* 'my girl'. I have never heard anybody use the phrases *poi horuŋ* 'my boy' or *kel enyau* 'my girl' in this way.

<sup>14</sup> This often reduces in fast speech to *naŋpoi* or even *nampoi*.

I have very little idea about the sort of interpretation which should be given to the distribution of this particular possessive form. One possibility is that given its frequent association with a pair of borrowed words, it may itself be a relatively recent borrowing. If this is derived from a language to the north, one possible recent locus for the transfer would be the sandalwood stations that operated with labourers from islands to the north between 1854–64 (Crowley 1990:60–63). That is, perhaps *najku* ‘my’ was part of the linguistic repertoire available to speakers of incipient Bislama. However, there is no documentary evidence in support of such a suggestion.

Interestingly, however, there is some external evidence to suggest that perhaps a first person possessive form might be more ‘borrowable’ than other elements of a paradigm. In recent years, there has been a number of holophrastic borrowings into Bislama of phrases of the type *my* + *NOUN* in English, as in *maefren* ‘my friend’, *maebrata* ‘my brother’, *maekraes* ‘my Christ’ (used as an interjection), *maetaem* ‘my time (to drink kava)’. No other possessive pronoun has been involved in borrowings of this type in Bislama. Bradshaw (pers. comm.) also points out that the English pronoun *my* has also been borrowed into Japanese as *mai*, with the meaning of ‘private’, as in *mai kaa* ‘private car’. Once again, no other possessive pronoun has been borrowed from English into Japanese.

## 7 Conclusion

There is much in §6 that has been extremely speculative. Given the nature of our sources, it is probably inevitable that we will not be able to say anything more definitive with relation to many of the points that I have raised. I hope, however, that my major point is seen as being tenable, namely that what happened to the morphological expression of possession in the languages of Erromango was probably due to a combination of factors.

The Erromangan languages had probably already innovated in the morphological expression of possession. By the Proto Erromangan stage, it seems that the distinction between edible, drinkable and general alienable possession had been lost. It also seems likely that a competing pattern of indirect possession had emerged in which nominal and pronominal possession were expressed in similar ways. Finally, Proto Erromangan also seems to have begun to allow independent pronouns to function as possessors in direct possessive constructions, in competition with the directly inherited pronominal suffixes.

It is also likely that there was some accidental—and possibly even deliberate—missionary tinkering with grammatical patterns which then became established in the language, if only as stylistic alternatives. At the same time, contact phenomena among Erromangans and between Erromangans and speakers of other languages were probably acting to promote certain patterns over others, and allowing for the spread of forms from one language to another. However, arguments that these languages have been structurally compromised as a result of colonial contact—as implied in Mühlhäusler (1996)—are not sustainable.

## References

- Crowley, Terry, 1990, *Beach-la-Mar to Bislama: the emergence of a national language in Vanuatu*. Oxford Studies in Language Contact. Oxford: Clarendon Press.
- 1995, How did Erromangan verbs get to be so messy? Paper presented at the Second International Conference in Oceanic Linguistics. Suva, Fiji, 3–7 July.
- 1997, What happened to Erromango's languages? *Journal of the Polynesian Society* 106(4):33–63.
- 1998a, *An Erromangan (Sye) grammar*. Oceanic Linguistics Special Publication No. 27. Honolulu: University of Hawai'i Press.
- 1998b, *Ura*. Languages of the World Materials, 240. München: LINCOM EUROPA.
- Dutton, Tom, 1985, *Hiri Motu iena sivarai*. Port Moresby: University of Papua New Guinea Press.
- Gordon, Rev. J.D., 1889, Sketch of the Erromangan grammar. In Rev. D. MacDonald, ed. *Three New Hebrides languages (Efatese, Erromangan, Santo)*, 61–84. Melbourne: Melbourne Public Library.
- Lynch, John, 1983a, The languages of Erromango. In John Lynch, ed. *Studies in the languages of Erromango*, 1–10. Canberra: Pacific Linguistics.
- 1983b, *Ura* grammar sketch and vocabulary. In John Lynch, ed. *Studies in the languages of Erromango*, 148–183. Canberra: Pacific Linguistics.
- 1995, Linguistic subgrouping in Vanuatu and New Caledonia: some preliminary hypotheses. Paper presented at the Second International Conference in Oceanic Linguistics. Suva, Fiji, 3–7 July.
- Lynch, John and A. Capell, 1983, *Sie* grammar outline. In John Lynch, ed. *Studies in the languages of Erromango*, 11–74. Canberra: Pacific Linguistics.
- Mühlhäusler, Peter, 1996, *Linguistic ecology: language change and linguistic imperialism in the Pacific region*. London: Routledge.
- Robertson, H.A., 1902, *Erromanga: the Martyr Isle*. New York: A.C. Armstrong and Son.

# *The Gilbertese -i intransitives, high vowel erasure and related phenomena*

SHELDON P. HARRISON

The suffix POC \*-i has a solid pedigree as a marker for Oceanic transitive verb forms, whose corresponding intransitive forms are generally unaffixed. Its PAN antecedent was likely to have been a valency-increasing device of some sort. But in some Micronesian languages, a suffix homophonous with the transitive *-i* appears, rather unexpectedly, as a marker of intransitive, as well as transitive verb forms. Most of the evidence for an intransitive *-i* comes from Gilbertese, but some evidence can also be adduced from the Ponapeic languages. A major problem in assessing putative *-i* intransitive forms in Gilbertese is distinguishing between verb roots with final *-i* and an *-i* intransitive suffix. This distinction is rendered opaque by a process of high-vowel erasure that makes many roots with historically final high vowel appear to be consonant final. When such forms are factored out, some twenty Gilbertese verb roots remain for which an intransitive suffix *-i* must be postulated.

## 1 Transitive–Intransitive alternations in recent linguistic theory<sup>1</sup>

Grammatical theorists have probably known for a very long time that the term **transitive verb** is a rather dangerous one, because many verbs that one thinks of as ‘semantically’ transitive can be used quite regularly in intransitive contexts. Perlmutter and Postal (1984) contributed the terms **unaccusative** and **unergative** to the grammatical lexicon to describe two such patterns of alternation in English, exemplified respectively by (1) and (2).

<sup>1</sup> It is an honour for me to offer this paper to my friend and teacher, Byron Bender. I only wish I had a better effort to offer for this volume. Thanks to Ken Rehg for comments and suggestions on an earlier draft. Language name abbreviations used herein are: Gil – Gilbertese, Kos – Kosraean, Mok – Mokilese, PAN – Proto Austronesian, PMc – Proto Micronesian, POC – Proto Oceanic, Pon – Ponapean (Pohpeian), Tru – Trukese (Chuukese), Wol – Woleaian. Other abbreviations: 1 – first person, 2 – second person, 3 – third person, ANIM – animate, ART – article, C – consonant, CN – common noun, CNST – construct suffix, INANIM – inanimate, NP – noun phrase, PL – plural, SG – singular, s.o. – someone, s.th. – something, V – vowel.

- (1) Hilda dropped the plate.  
The plate dropped.
- (2) Fred ate the banana.  
Fred ate.

Such alternations have become a focus of attention in 'mainstream' generative theory in the last decade, since the publication of Burzio's (1986) results concerning intransitivity in Italian. **Conative** alternations like (3) are less often noted in syntax texts, possibly because the 'valency' of the verb doesn't change in such cases. To this list of alternation types we should probably also add **object incorporation**, as in (4), although in English in most cases the intransitive alternate is not syntactically verbal.

- (3) Hilda shot Fred.  
Hilda shot at Fred.
- (4) We're frying fish.  
We're having a fish fry.

Unaccusative, unergative and conative alternations have also received attention in work on verbal aspect and, in particular, by scholars like Pustejovsky (1995) and Levin (1993) in the context of the structure of the lexicon. The latter has developed a rich syntactico-semantic typology of transitive-intransitive alternations for English: four subclasses of unaccusatives, eight of unergatives, and (by my reckoning at least) three of conatives.

## 2 Transitive–Intransitive alternations in Oceanic

Oceanists have been dealing with transitive–intransitive alternations since long before the subject became fashionable. We even had our own terms for unaccusative and unergative: Pawley (1973) (following Clark; see Clark 1976) termed them **P-verbs** (patient-oriented) and **A-verbs** (agent-oriented); and Chung (1978) used the term **middle** to refer to the intransitive member of a conative-type alternation in Polynesian.

The salience of transitive–intransitive alternations for Oceanists is probably a consequence of the fact that the alternations have morphological, as well as syntactic consequences. In phonologically prototypical Oceanic languages (in which Proto Austronesian and Proto Oceanic final vowels are preserved, but final consonants are lost), intransitive forms have no particular morphological marking, while related transitive forms carry a reflex of one or the other of Proto Oceanic **primary** transitive suffix \*-i or **associative** transitive suffix \*-aki[ni]. If the verb root was one that was consonant-final in Proto Oceanic, then the transitive form will typically reflect the historical root-final consonant as a so-called **thematic** consonant<sup>2</sup>, which is lost in the unsuffixed intransitives, as in the derivation in (5) from pre-Gilbertese forms for 'to climb':

- |     |         |                        |  |              |            |
|-----|---------|------------------------|--|--------------|------------|
| (5) | pre-GIL | *tamwak+a              |  | *tamwak      |            |
|     |         | _____                  |  | *tamwa       | C-deletion |
|     | Gil     | <i>tamwaka</i>         |  | <i>tamwa</i> |            |
|     |         | 'to climb (for) s.th.' |  | 'to climb'   |            |

<sup>2</sup> The fact that the thematic consonant is often not etymological has been widely reported.

In less typical languages, in which Proto Austronesian and Proto Oceanic final vowels are also lost, there is no overt synchronic reflex of the primary transitive suffix.<sup>3</sup> Given a historical short vowel transitive suffix, the transitive form in such languages reflects the Proto Austronesian and Proto Oceanic verb root, while the intransitive form is that root minus its final syllable, thus the derivations in (6) for PMC<sup>4</sup> verb root \*wadek 'to calculate, count, read':

(6)		GILBERTESE		PONAPEAN		
	pre-PMC	*wadek+a	*wadek	*wadek+a	*wadek	
		_____	*wade	_____	*wade	C-deletion
		_____	_____	*wadek	*wad	V-deletion
		<i>wareka</i>	<i>ware</i>	<i>wadek</i>	<i>wad</i>	

In some languages, standard Fijian being one that comes immediately to mind, the primary transitive suffix *\*-i* is directly reflected only with pronominal and proper noun objects; with common noun objects the suffix is *-a*.

(7)	STANDARD FIJIAN	
	<i>Au a raica na gone.</i>	'I saw the boy.'
	<i>Au a raici Paula.</i>	'I saw Paul.'
	<i>Au a raici koya.</i>	'I saw him.'

The standard account of this variation is that the *-a* form is an irregular reflex of an earlier bimorphemic *\*-ia*, composed of the *\*-i* transitive suffix and a third person singular object pronominal suffix. Forms like *raicia* are reflected in other Fijian languages. In standard Fijian, the standard account claims, there has been irregular elision/deletion of the *\*i* in the common noun object forms, leaving only the third person singular pronominal suffix.

I've never believed the standard account of Fijian *-a*, for what I think are two very good reasons. First, irregular morpheme-specific phonological change does not really provide much of an account of anything, since one can always appeal to such changes without fear of refutation. If there are two accounts of some phenomenon, one that requires an irregular phonological change and one that does not, then (all other things being equal) the latter is a better account. Second, there is considerable evidence in Eastern Oceanic (and farther afield in Austronesian) for a verbal suffix *-a* that is not clearly identifiable with the third singular object suffix. It is found in Rotuman, in Polynesian languages (where it alternates with reflexes of *\*-ia*) and in the Micronesian languages, with which I am most familiar.

Consider, for example, the full paradigms for the Gilbertese verbs *wareka* 'to calculate, to read s.th.' and *wetea* 'to call s.o.' (exemplifying the paradigms for consonant- and vowel-final roots, respectively), shown in (8). In Gilbertese, as in standard Fijian, the suffix *-i* appears before all pronominal objects except the first person singular.<sup>5</sup> The suffix *-a* is used before any singular noun phrase object. In the third person plural, Gilbertese contrasts an animate

<sup>3</sup> I choose this convoluted wording because languages like Mokilese and Ponapean, which have lost final vowels, often do have a productive transitive suffix. In Mokilese, it is *-i* and in Ponapean *-ih* (e.g. *Mok loakjid* 'to fish', *loakjidi* 'to fish for s.th.'). One infers that such transitives arose after the loss of final vowels or were otherwise immune from it.

<sup>4</sup> Proto Micronesian reconstructions follow Jackson (1983).

<sup>5</sup> In the first person singular, one would expect the nonoccurring *\*warekiai*. Such forms are reflected in other Micronesian languages, but not in Gilbertese. Other evidence lacking, I must postulate irregular phonological change in these forms in Gilbertese! (Yes, even I occasionally take the coward's route!)



form with an object suffix reflecting PMc \*-ira '3P' and an inanimate form that may reflect POc/PMc \*-i transitivising suffix with no additional object suffix.

(8)	SINGULAR	PLURAL	SINGULAR	PLURAL	
1	<i>warekai</i>	<i>warekiira</i>	<i>weteai</i>	<i>weteiira</i>	
2	<i>warekiko</i>	<i>warekingkamii</i>	<i>weteiko</i>	<i>weteingkamii</i>	
3	<i>warekia</i>	<i>warekiia</i>	<i>weteia</i>	<i>weteiia</i>	ANIMATE
NP	<i>wareka</i>	<i>wareki[i]</i> <sup>6</sup>	<i>wetea</i>	<i>wetei</i>	INANIMATE

Transitive forms reflecting PMc \*-a are also found in western Trukic languages, though with a function that is not entirely clear. Since in other Micronesian languages short final vowels are lost, any other evidence for \*-a transitives must be indirect. Such evidence can be found in patterns of morphophonemic alternation involving verbs with historical \*a.

Let me first motivate a process, in the Ponapeic languages, of assimilatory raising of \*a before a high vowel in a following suffix. In that regard, consider the possessive paradigms in (9) for the noun reflecting PMc \*tama 'father' in Gilbertese and in Mokilese. (The Mokilese plural forms cited are those cognate with the Gilbertese plurals, though they have somewhat different semantics in Mokilese. The usual Mokilese plurals reflect older trials, and are a Ponapeic innovation. Note also that Gilbertese is unusual in having lost all first person exclusive forms, and for that reason Mok *jememwi* '1EXCLUSIVE' has been omitted from the paradigm in (9). The suffix-possessed paradigms in Micronesian languages have a construct form, reflecting 'capture' of the genitive preposition \*ni and used in construction with nonpronominal possessors. In Gilbertese, the third plural form is used with plural animate noun phrase possessors, however.)

(9)	GILBERTESE		MOKILESE	
	SINGULAR	PLURAL	SINGULAR	PLURAL
1	<i>tamau</i>	<i>tamara</i>	<i>joamoai</i>	<i>jamahs</i>
2	<i>tamamw</i>	<i>tamamii</i>	<i>joamoamw</i>	<i>jememi</i>
3	<i>tamana</i>	<i>tamaia</i>	<i>jamah</i> <sup>7</sup>	<i>jamahr</i>
CNST	<i>taman</i>		<i>jemen</i>	

Gilbertese does not reflect historical high vowels after nasals; the reconstructed forms of the second person singular and the construct suffix are PMc \*-mu and PMc \*-ni, respectively. Observe that, in Mokilese, when the historical final vowel of the suffix was nonlow, the short low vowel of the stem was raised;<sup>8</sup> when the historical final vowel of the suffix was low, no raising took place.

Now consider the transitive–intransitive pairs in (10), likely cognates in Gilbertese and Mokilese. (In this and subsequent lists, the intransitive form appears above the corresponding transitive.)

<sup>6</sup> The final vowel of the inanimate plural may be either long or short after consonant-final roots. This free variation still begs an account. So far as I am aware, with vowel-final roots the suffix is never long.

<sup>7</sup> The third person singular possessive suffix has been reconstructed as PMc \*-ña. PMc \*ñ is regularly lost in Ponapeic languages (though in Mokilese, conditions apply). Why the resulting final vowel in this class of Mokilese third person singular possessives is long has always been a mystery to me.

<sup>8</sup> Other factors determine the resulting height and backness of the originally low vowel.



(10)	GILBERTESE <sup>9</sup>	MOKILESE	
a.	<i>ware</i>	<i>wadwad</i>	'count'
	<i>wareka</i>	<i>wadek</i>	
	<i>beka</i>	<i>poak</i>	'defecate'
	<i>bekara</i>	<i>pakad</i>	
	<i>mumuta</i>	<i>umwwuj</i>	'vomit'
	<i>mumutaa</i>	<i>umwwujoa</i>	
	<i>urung</i>	<i>widing</i>	'steer' (Mok 'to string s.o. along')
	<i>urunga</i>	<i>widinge</i>	
b.	<i>taim</i>	<i>joaijoai</i>	'sharpen'
	<i>taima</i>	<i>jaim</i>	
	<i>taun</i>	<i>doau</i>	'bury'
	<i>tauna</i>	<i>daun</i>	
		<i>doau</i>	'climb'
		<i>daur</i>	
	<i>ati</i>	<i>poaipoi</i>	'weave'
	<i>ata</i>	<i>pa</i>	

As evidenced in (10b), it is generally the transitive forms in Mokilese, rather than the intransitives, that reflect a historical low vowel as low. This is precisely what one would expect were the deleted transitive suffix \*-a and not \*-i. The only cases in which a historical \*a in the root is reflected as nonlow in Mokilese transitive forms are those in which that vowel was root-final, as in *umwwujoa* 'to vomit on s.th.'. (At some earlier period, short final *a* was proscribed in Mokilese; the only cases of which I am aware are those in which a final consonant was more recently lost, as in *pa* 'to weave s.th.'). Those cases aside, if there is a raised reflex of an earlier \*a, it is in fact the **intransitive** form of the relevant lexeme that displays it.

### 3 Micronesian intransitive forms in -i

For 'typical' Oceanic languages, the morphological difference between transitive and intransitive forms of the same verb lexeme is that the former take a vocalic suffix that the latter do not. We would therefore expect the transitive form to be longer than the corresponding intransitive, by derivations like those in (11):

<sup>9</sup> The *e* in Gilbertese *beka* 'defecate' is unexpected on comparative evidence. The correspondence Mok *d* : Gil *t* in the forms for 'bury' is also irregular.

(11)	GILBERTESE		MOKILESE		
	INTRANSITIVE	TRANSITIVE	INTRANSITIVE	TRANSITIVE	
PMc	*wadek	*wadek+a	*wadek	*wadek+a	
	<i>wade</i>	_____	<i>wade</i>	_____	C# > Ø
	_____	_____	<i>wad</i>	<i>wadek</i>	V# > Ø
	<i>ware</i>	<i>wareka</i>	<i>wad</i>	<i>wadek</i>	other changes
PMc	*mumuta	*mumuta+a	*mumuta	*mumuta+a	
	_____	_____	_____	_____	C# > Ø
	_____	_____	<i>mumut</i>	<i>mumuta</i>	V# > Ø
	<i>mumuta</i>	<i>mumutaa</i>	<i>umwwuj</i>	<i>umwwujoa</i>	other changes
PMc			*taim	*taim+a	
			<i>tai</i>	_____	C# > Ø
			_____	<i>taim</i>	V# > Ø
			<i>joai</i>	<i>jaim</i>	other changes

What we do not expect are the cases in boldface in (12), in which, after any vowels lost through final-vowel deletion<sup>10</sup> are restored, transitive and intransitive forms are the same length. The only possible explanation for such cases, consistent with other facts of the languages in question, is that the transitive is not formed by adding a vocalic suffix to the intransitive, but that both transitive and intransitive forms are/were vowel-final.

(12)	GILBERTESE	MOKILESE	PONAPEAN	
	<i>taim</i>	<i>joaijoai</i>	<i>seisei</i>	'to sharpen'
	<i>taima</i>	<i>jaim</i>	<i>saim</i>	
	<i>ati</i>	<i>poai</i>		'to weave'
	<i>ata</i>	<i>pa</i>		
	<i>aang</i>	<i>koarorang</i>	<i>kereng</i>	'to warm/dry over a fire'
	<i>aanga</i>	<i>karang</i>	<i>karang</i>	

I will refer to the intransitive forms in question as '-i intransitives', since the final vowel of the intransitive is usually /i/. This is transparently the case in Gilbertese, where the forms either end in *i* (rarely *u*; see §4 below) or a nasal consonant. Since final nasals in Gilbertese arise through loss of a final high vowel, nasal-final intransitives can be regarded as having a final high vowel historically. The patterns of morphophonemic alternation in relevant transitive-intransitive pairs in the Ponapeic languages also suggest a historical final \*i, since intransitive forms show mid vowels alternating with low vowels in the corresponding transitives. Thus, I propose derivations like (13) (for the Ponapean verb 'to dry over a fire').

(13)	PMc	*karangi	*karanga	
		<i>kerengi</i>	_____	umlaut (raising of *a before final *i)
	Pon	<i>kereng</i>	<i>karang</i>	V# > Ø

I would suggest that at least three of Rehg's (1981) transitive verb classes in Ponapean have -i intransitives. The most transparent, those Rehg terms **ablaut intransitives**, show the morphophonemic alternations expected of historical \*i-final forms, as in (14). To this set I

<sup>10</sup> As noted earlier, Gilbertese loses historical final high vowels after nasal consonants. The digraph *oa* in Mokilese represents /ɔ/. I assume pre-Mokilese forms \*pai and \*paa 'to weave', in which the final glide of the former remains, but the final long vowel of the latter is short. These changes are regular in Mokilese.

would also add the examples in (15). The intransitive form Pon *keng* is used only in incorporated-object constructions, as in *kengwini* 'to take (lit. eat) medicine'; the verb *mwenge* (Mok *mwinge*) serves as the 'suppressed object' form.<sup>11</sup>

- |      |                |                 |                       |
|------|----------------|-----------------|-----------------------|
| (14) | PONAPEAN       | MOKILESE        |                       |
|      | <i>peleng</i>  | <i>poaloang</i> | 'spread out to dry'   |
|      | <i>palang</i>  | <i>palang</i>   |                       |
|      | <i>kereng</i>  | <i>koaroang</i> | 'dry over a fire'     |
|      | <i>karang</i>  | <i>karang</i>   |                       |
|      | <i>periper</i> |                 | 'to cut'              |
|      | <i>par</i>     |                 |                       |
|      | <i>epid</i>    |                 | 'carry on one's side' |
|      | <i>apid</i>    | <i>apid</i>     |                       |
|      | <i>der</i>     |                 | 'to strike (of fish)' |
|      | <i>dar</i>     |                 |                       |
|      | <i>ngked</i>   | <i>ingkoad</i>  | 'to roof with thatch' |
|      | <i>ngkad</i>   |                 |                       |
|      | <i>lamai</i>   |                 | 'to be cruel'         |
|      | <i>leme</i>    |                 |                       |
|      | <i>keikei</i>  | <i>keikei</i>   | 'to bite'             |
|      | <i>ke</i>      | <i>koa</i>      |                       |
| (15) | <i>keng</i>    | <i>kang</i>     | 'to eat'              |
|      | <i>kang</i>    | <i>kang</i>     |                       |

In Reh's **reciprocal intransitives** (16) and **reduplicated intransitives** (17), the transitive and intransitive forms are identical in respect to the reciprocal prefix, reduplication and short, low-vowel raising in the form *pekekil*. This fact can be accounted for by analysing them as *-i* intransitive. Most do not have short, low root vowels of the sort expected to result in the morphophonemics seen in the ablaut intransitives. Of those that do, *kakil* 'to stare' does show the alternation diagnostic of *-i* intransitives; *ese* 'to know', from PMc \**ata* (Gil *ataa* 'to know s.th.' *te tia atai bwai* 'a know-it-all'), shows a mid vowel in both transitive and intransitive forms. For those Ponapean reduplicated intransitives that give evidence of the final vowel, that vowel is high, just as expected.

- |      |                |              |                                |
|------|----------------|--------------|--------------------------------|
| (16) | PONAPEAN       |              |                                |
|      | INTRANSITIVE   | TRANSITIVE   |                                |
|      | <i>pidilin</i> | <i>dilin</i> | 'to pull [one another's] hair' |
|      | <i>pekekil</i> | <i>kakil</i> | 'to stare at [one another]'    |
|      | <i>pehse</i>   | <i>ese</i>   | 'to know [one another]'        |
|      | <i>pisiken</i> | <i>siken</i> | 'to push [one another]'        |
|      | <i>paiahn</i>  | <i>ahn</i>   | 'to be used to [one another]'  |

<sup>11</sup> By 'suppressed object form' I mean a form with an argument (in thematic structure or what have you) that would be expected to be realised syntactically as object, but is lexically barred from any overt realisation. We would expect the nonoccurring form \**koang* in Mokilese as the incorporated object form. In Gilbertese, the transitive forms for 'to eat', with one exception, have the stem *kan-*: *kaniko* 'to eat you', *kana* 'to eat s.th.'. The exception in Gilbertese is the third person plural inanimate, which is *kang*. That same form is also used in incorporated object constructions.

## (17) PONAPEAN

INTRANSITIVE	TRANSITIVE	
<i>eliel</i>	<i>el</i>	'massage'
<i>pilipil</i>	<i>pil</i>	'choose'
<i>popohr</i>	<i>pohr</i>	'slap in anger'
<i>pipihs</i>	<i>pihs</i>	'urinate'
<i>usuhs</i>	<i>us</i>	'pull out'

**4 Gilbertese intransitives in -i**

In my earliest musings on Micronesian verbs with intransitives in \*-i (Harrison 1978:1071ff.), I was concerned with the transitive verb forms, not their intransitive counterparts, and I passed over the latter without comment. I also knew very little about Gilbertese at that time. Subsequent fieldwork in Kiribati, and more careful study of published material on Gilbertese, has revealed a substantial class of -i intransitive verbs in that language. My current notes list upwards of sixty verbal lexemes with intransitive forms in -i (including nasal-final intransitives, which, as noted above, point to a deleted final high vowel). These intransitive (or, perhaps more accurately, nontransitive) forms are variously used:

- (i) as intransitive verbs
- (ii) as action or result nominalisations
- (iii) in incorporated object constructions

In uses (i) and (ii) they are frequently reduplicated. These various uses are exemplified with both 'regular' intransitives, as in (18), and -i intransitives, as in (19). Note that some roots have both regular and -i intransitives. These lists are representative, not exhaustive.

## (18) GILBERTESE REGULAR INTRANSITIVES

<i>tamwaka</i>	'to climb on s.th.'
<i>tamwatamwa</i>	'to climb'
<i>te tamwa</i>	'a (way of) climb(ing)'
<i>minota</i>	'to twist s.th.'
<i>te mino</i>	'a twist, turn'
<i>minomino</i>	'to be twisted'
<i>korea</i>	'to cut s.th.'
<i>te koro</i>	'a cut'
<i>koro karewe</i>	'to cut toddy'

## (19) GILBERTESE INTRANSITIVES IN -i

<i>unika</i>	'to plant s.th.'
<i>te unuuniki</i>	'the/an act of planting shoots'
<i>ongira</i>	'to squeeze/wring out s.th.'
<i>ongngong</i>	'spongy, a sponge'
<i>ongiri ngako</i>	'to blow the nose'
<i>ikota</i>	'to add/gather up'
<i>iko ben</i>	'to gather coconuts'

<i>te ikoti</i>	‘sum, account’
<i>ikoti mwane</i>	‘to prepare an account’ (not 3PL because <i>mwane</i> is singular)
<i>koroma</i>	‘to husk (a coconut)’
<i>te koro</i>	‘husking stick’
<i>te korokoro</i>	‘act of husking’
<i>korom ben</i>	‘to husk coconuts’
<i>kinika</i>	‘to pinch/pluck s.th.’
<i>kin uti</i>	‘to pick out lice’
<i>kiniki uee</i>	‘to pick flowers’
<i>ataa</i>	‘to know s.th.’
<i>te atai</i>	‘knowledge’
<i>bita</i>	‘to exchange/invert s.th.’
<i>ibiibiti</i>	‘to change places’
<i>buoka</i>	‘to help s.o.’
<i>ibuobuoki</i>	‘to collaborate’
<i>kimra</i>	‘to pack/compact s.th.’
<i>kimri</i>	‘to be packed tight’
<i>rabwata</i>	‘to embrace s.o.’
<i>irabwarabwati</i>	‘to embrace one another’

This more extensive body of data makes it possible to draw firmer conclusions on the morphological status of the final *-i*. The specific synchronic question I want to answer is whether the *-i* termination of these intransitives is:

- (i) part of the verb root, or
- (ii) a hitherto unrecognised intransitive suffix.

For hypothesis (i) to hold, it must be the case that the morphological structure of transitive forms from such *i*-final roots is **not** agglutinative. That possibility goes against the usual assumption for transitive morphology in prototypical Oceanic languages. With agglutinative transitive morphology, removing any transitive affixation from a transitive verb form should reveal the verb root. Consider the Gilbertese transitive verb form *weteiko* ‘to call you (SG)’ [= *wete+i+ko*]. Removing the object suffix and the transitive suffix from that form yields the root  $\sqrt{wete}$  ‘to call’. Removing the transitive suffix from *tuanga* ‘to tell s.o.’ or *teboka* ‘to wash s.o.’ reveals the roots  $\sqrt{tuang}$  ‘tell, order’ and  $\sqrt{tebok}$  ‘wash’, respectively. Removing the transitive morphology from transitive verb forms like *ira* ‘to follow s.o.’ or *iriira* ‘to follow us’ yields a root  $\sqrt{ir}$  ‘follow’. But the leftmost *r* of forms of this verb is not a thematic consonant, appearing only in suffixed transitive forms; that is, the intransitive form is not \**i* or \**ii*, but the *-i* intransitive *iri* ‘to follow’. If the final *-i* is part of a root  $\sqrt{iri}$ , then the transitive morphology is not agglutinative, since removing that morphology does not yield the root.

It follows, then, that if transitive morphology is agglutinative, then the *-i* of the intransitive *iri* ‘to follow’ must be a suffix. But if the transitive morphology for *-i* intransitive verbs is **not** agglutinative in the sense just outlined, then other conclusions are possible. For example, if one allows the possibility that the word-formation process deriving transitive verb forms

erases a (high) root-final vowel before adding transitive suffixes, then an intransitive form in *-i* can be identified with the verb root and the transitive stem.

## 5 High-vowel-final roots and erasure in Gilbertese

The transitive morphology of Gilbertese roots with final nonhigh vowels was exemplified above in the paradigm for *wetea* ‘to call s.o.’ in (8). Regular roots ending in short, high vowels have an identical paradigm, as in (20) for *otorikia* ‘to smash s.th.’ and *taua* ‘to hold s.th.’.

(20)	SINGULAR	PLURAL	SINGULAR	PLURAL	
1	<i>otorikiai</i>	<i>otorikiira</i>	<i>tauai</i>	<i>tauira</i>	
2	<i>otorikiiko</i>	<i>otorikingkamii</i> <sup>12</sup>	<i>tauuko</i>	<i>tauingkamii</i> <sup>13</sup>	
3	<i>otorikiia</i>	<i>otorikiia</i>	<i>tauua</i>	<i>tauia</i>	ANIMATE
NP	<i>otorikia</i>	<i>otorikii</i>	<i>taua</i>	<i>tauu</i>	INANIMATE

Durrheimer (1964) used the term “strong” to contrast these paradigms, along with those of *wareka* ‘to calculate s.th.’ and *wetea* ‘to call s.o.’, with another vowel-final root paradigm, the ‘weak’ paradigm, in which the *-i-* normally preceding a pronominal suffix is absent **in the singular**. In such paradigms, 3SG and NPSG forms are identical, as in the paradigms in (21) for *mwiiokoa* ‘to entrust to s.o.’ and *ikua* ‘to pound s.th.’.

(21)	SINGULAR	PLURAL	SINGULAR	PLURAL	
1	<i>mwiiokoai</i>	<i>mwiiokoiira</i>	<i>ikuai</i> <sup>14</sup>	<i>ikuira</i>	
2	<i>mwiiokoko</i>	<i>mwiiokoingkamii</i>	<i>ikuko</i>	<i>ikuingkamii</i>	
3	<i>mwiiokoa</i>	<i>mwiiokoiia</i>	<i>ikua</i>	<i>ikuia</i>	ANIMATE
NP	<i>mwiiokoa</i>	<i>mwiiokoi</i>	<i>ikua</i>	<i>ikui</i>	INANIMATE

The weak paradigm is particularly common with causative<sup>15</sup> verbs derived via the prefix *ka-*, as in *kabaea* ‘to tie it/s.th.’ (< *bae* ‘tied, obligated’). Whether a vowel-final root has a weak paradigm like *mwiiokoa*, or a strong one like *wetea*, is a lexical matter. And for roots with a final short, high vowel in particular, speakers are often undecided regarding which model to follow. For example, *otorikia* and *ikua* have alternative weak and strong paradigms, respectively, with singular forms (22).

<sup>12</sup> also *otorikiingkamii*

<sup>13</sup> also *tauingkamii*

<sup>14</sup> The root-final /u/ often assimilates to the preceding /i/ in these forms, yielding forms like *ikiiai* ‘to pound me’ and *ikiira* ‘to pound us’. As this last form demonstrates, overlong vowels are not permitted in Gilbertese.

<sup>15</sup> The label ‘causative’ for Gilbertese *ka-* and its cognates in other Micronesian languages is more appropriate than is often the case for so-called ‘causative’ morphology. The majority of forms in *ka-* have CAUSE semantics. They do not all have transitive syntax, however. Nontransitive forms in *ka-* typically have what Levin (1993:39) describes as “characteristic property of agent” semantics: as in Gil *kakamaaku* ‘frightening’ < *maaku* ‘to be afraid’; *te kaetieti* ‘announcement’ < *eti* ‘to be straight, correct’; *kakamate* ‘dangerous, deadly’ < *mate* ‘to die’.

(22)	WEAK SINGULAR	STRONG SINGULAR
	1 <i>otorikiai</i>	<i>ikuai</i>
	2 <i>otorikiko</i>	<i>ikuiko</i> (~ <i>ukuuko</i> )
	3 <i>otorikia</i>	<i>ikuia</i> (~ <i>ukuua</i> )
	NP <i>otorikia</i>	<i>ikua</i>

Note that what characterises the weak vowel-final paradigm is just the absence of the *-i*-transitive formative in the singular. This phenomenon need not, and probably should not, be described as an erasure. Since *-i-* is a morpheme, the weak paradigm is adequately described by simply saying that in the singular, that morpheme is just not added.

But the paradigm confusion for Gilbertese verbs which end in a short, high vowel is not limited to indecision regarding the choice between a weak and a strong paradigm. Speakers sometimes appear to take the matter one step further, and give these verbs a strong thematic (i.e. consonant-final) paradigm. I will call that paradigm the ‘mixed’ paradigm, as exemplified in (23).

(23)	SINGULAR	PLURAL	SINGULAR	PLURAL	
	1 <b><i>otorikai</i></b>	<i>otorikiira</i>	<b><i>ikai</i></b>	<i>ikuiira</i>	
	2 <i>otorikiko</i>	<i>otorikingkamii</i>	<i>ikiko</i>	<i>ikuingkamii</i>	
	3 <i>otorikia</i>	<i>otorikiia</i>	<i>ikia</i>	<i>ikuiia</i>	ANIMATE
	NP <b><i>otorika</i></b>	<i>otorikii</i>	<b><i>ika</i></b>	<i>ikui</i>	INANIMATE

Ignoring high-vowel assimilation, the members of these two paradigms that are neither vowel-final weak nor vowel-final strong are the first and third person singulars.<sup>16</sup> In those forms, shown in bold in (23) and (24), it is not the transitive formative that is missing. Rather, the final short, high vowel of the root has been erased. At least one verb, *maaka* ‘to fear s.th.’ (*maaku* ‘to be afraid’), has only a thematic paradigm in the singular at least (24).

(24)	SINGULAR	PLURAL
	1 <b><i>maakai</i></b>	<i>maakuira</i>
	2 <i>maakuko</i>	<i>maakingkamii</i>
	3 <i>maakua</i>	<i>maakiia</i> ~ <i>maakuiia</i>
	NP <b><i>maaka</i></b>	<i>maakii</i>

Its causative *kamaakua* (~ *kamaaka*) ‘to frighten s.o.’ has the three competing paradigms. Exclusive mixed paradigms are particularly common in causatives derived from final high vowel (or nasal) nontransitives, however. Those in my data with an attested final *-i* nontransitive form (intransitive verb or common noun without *ka-*) include:

<sup>16</sup> The invariant passive in *-aki*. Passives have not been shown in any of the example paradigms, but are properly part of those paradigms. For *otorikia* and *ikua*, the passives in both weak and strong paradigms are *otorikiaki* and *ikuaki* (~*ikiaki*), respectively. In the ‘deviant’ thematic paradigms, the passives are *otorikaki* and *ikaki*, respectively.

(25) *ka*-TRANSITIVE VERB<sub>NP</sub>

<i>kabanga</i>	'make fly by'
<i>kabata</i>	'augment'
<i>kabeta</i>	'float'
<i>kabinota</i>	'infect, spread smell'
<i>kabura</i>	'cause pus'
<i>kabwaka</i>	'heap up'
<i>kabwaraaka</i>	'overturn'

*kaeta* 'straighten'

<i>kaika</i>	'bail, drain'
<i>kaikawaa</i>	'raise, bring up'
<i>kaira</i>	'lead'
<i>kamooa</i>	'make/let drink'
<i>kamota</i>	'cut off, end'
<i>kamwaa</i>	'cook'
<i>kamwaita</i>	'increase'
<i>kaota</i>	'show'
<i>karaita</i>	'dazzle'
<i>karaoa</i>	'do, make'
<i>karika</i>	'conceive'

*katea* 'build'

*kauka* 'open'

## INTRANSITIVE VERBS OR COMMON NOUN

<i>bang</i>	'fly by quickly'
<i>bati</i>	'[too] much'
<i>beibeti</i>	'float, drift'
<i>binoti</i>	'stench'
<i>buri</i>	'pus'
<i>bwaki</i>	'heap'
<i>bwaraaaki</i>	'upside down'
<i>kabwaraaki</i>	'fisherman's hat (made from coconut frond)'
<i>eti</i>	'straight, correct'
<i>kaetieti</i>	'announcement'
<i>iki</i>	'drained, dried up'
<i>ikawai</i>	'mature'
<i>iri</i>	'follow'
<i>mooi</i>	'drink'
<i>moti</i>	'piece, ended'
<i>mwai</i>	'cooked'
<i>mwaiti</i>	'numerous'
<i>oti</i>	'clear, manifest'
<i>raititi</i>	'shine, scintillate'
<i>raoi</i>	'good [result]'
<i>riki</i>	'become'
<i>kariki buaka</i>	'have a miscarriage (cf. <i>buaka</i> 'bad')'
<i>tei</i>	'stand'
<i>katei</i>	'custom, edifice'
<i>uki</i>	'opened'

There is little doubt that erasure, rather than a high-vowel intransitive suffix, is the source of the mixed paradigms of many of these causatives, and also of the noncausative transitives noted earlier. The relevant evidence is of three sorts:

## (i) evidence from etymology

Gilbertese *maaku* 'afraid' (*maaka* 'to fear s.th.') is a reflex of the well-supported PMc \**mataku*, an etymon that goes back to Proto Austronesian. (The loss of \**t* is a consequence of a regular constraint barring tautomorphemic apical and velar stops in native items.) PMc \**peti* 'float' is reflected as Kos *puhs*, Pon *pei*, Wol *pesi*, and is undoubtedly related to Proto Polynesian \**pati* 'float'.

At least three V[-hi]V[+hi]# roots have external cognates. One is  $\sqrt{atai}$  'to know': *atai* 'knowing, knowledge', *ataa* 'to know s.th.'. This root has numerous cognates, both inside and outside Micronesia, and has been reconstructed as PAN \**tahu* 'know'. It is interesting to note that its Micronesian cognates (Kos *etuh*, Pon *ese* 'to know s.th.', Mok *jehjoa* 'to not know s.th.', the latter with the negative prefix Mok *ja-*) have no 'true' intransitive counterpart in their respective languages. (Ponapean has the reciprocal *pehse* 'to know one another'). The second Vi# root with Micronesian cognates is *koi* 'to grate', *koa* 'to grate s.th.': Pon *koroiroi*, Wol *geriuger* 'to grind',



Pon *korehd*, Wol *geriu* 'to grate s.th.' 'to grind' 'to grind s.th.', Tru *keri-* 'grate' (a bound form appearing only in compounds). The third is *raoi* 'good', *karaoa* 'to create/do s.th.', which has been identified (Harrison 1994:338) as a Polynesian borrowing (cf. Sikaiana *laaoi* 'good').

(ii) evidence from morphology

At least four of the roots of mixed paradigm transitives are morphologically complex: *otoriki* 'smashed', *beibeti* 'float, drift', *raititi* 'shine' and *bwaraaki* 'overturned, upside down'. The first is a combination (a compound or a word+enclitic, depending upon how the second component is analysed) of *oto* 'broken' (*otea* 'to break s.th.') and *riki* 'much'. Its root-final *i* is the final vowel of *riki*. The next two cases are reduplications. The first of these apparently reflects an anomalous loss of **one** instance of *t*, while the second may also involve the prefix *ra-* 'somewhat, rather, a little' (as in *rabete* 'rather light' < *bete* 'light, not heavy', *rabeti* 'a little buoyant' < *beibeti* 'float'). If that is so, then it is probably derived from *iti* 'lightening'.

Gilbertese *bwaraaki* 'overturned, upside down' appears to be a passive, with the regular passive suffix *-aki*.<sup>17</sup> The corresponding transitive I only know as *bwaraa* 'put something in a *bwara*' (< *bwara* 'type of coconut leaf basket'). This etymology seems less farfetched when one understands that a *bwara* is a temporary basket woven quickly from a whole frond. When the weaving is complete, the frond is turned upside down and split along the rachis to form the basket. If *bwaraaki* is indeed a passive, then its final *i* is not a suffix *-i*. Nor is that of *moti* 'cut off, ended', which has a thematic transitive *motika* 'to cut' in addition to the causative transitive. Its final *i* could not conceivably be both a suffix and a reflex of the historical penultimate segment of the root.

(iii) evidence from morpheme structure

Since Dempwolff it has been recognised that most Austronesian major category roots were bivocalic objects of the form \*[C]V[C]V[C]. With the loss of word-final consonants in most Oceanic (including Micronesian) languages, these roots are reflected as [C]V[C]V. PAn \*[C]V[C] were rare under the usual interpretations of Proto Austronesian morpheme structure constraints. With this historical background, we expect Gilbertese forms with a suffix *-i* to have the shape [C]V[C]V[C]+*i*, for example *ata+i*, *kinik+i*, *korom+[i]*. We do not expect *te+i*, *uk+i*, *bet+i*, which have the syllable structure/canonical shape of roots, rather than affixed forms.

Under a synchronic erasure analysis, a root-final short, high vowel is erased before the addition of the suffixes *-ai*, *-a*, *-aki* or the pronominal object transitive formative *-i*. Erasure of this sort is not a common morphological phenomenon, and has not, to my knowledge, been postulated in any synchronic analyses of Austronesian morphology. Therefore, its occurrence in Gilbertese begs some sort of account. Given that it is not a general phonological process, but one restricted to short, high vowels in some verb forms, the most obvious account is

<sup>17</sup> When he first learned the title of this paper, Joel Bradshaw asked me whether there were also intransitive reflexes of POc \*-aki(ni) transitive suffix in Micronesian languages. The answer, of course, is yes; PMc agentless passive suffix \*-aki (which has developed into a full passive in Gilbertese and Kosraean) has without doubt the same historical source as the reconstructed POc transitive suffix. I have already treated this issue at great length in Harrison (1982).

analogical—erasure applies (in some cases perhaps redundantly) to generate paradigm members conforming to a thematic-stem paradigm in a paradigm that would otherwise conform to that of a vowel-final root. Like most analogical accounts, this one will not satisfy some readers. And it still begs at least one question; why is it restricted to roots with a final high vowel? One possible answer to that question is phonetic. High vowels are the least sonorant and, hence, the most likely to be lost, as evidenced elsewhere in the phonology of Gilbertese. But another factor may be the existence of an intransitive suffix *-i*.

## 6 The intransitive suffix *-i*

I will conclude that the final *-i* of an intransitive verb form in Gilbertese is a suffix, and not a root-final vowel, if it can be demonstrated internally or comparatively,

- (i) in the case of *Vi#* forms, that there are related forms in *V#*; and
- (ii) in the case of *Ci#* forms, that the consonant *C* is 'thematic'; that is, absent in homomorphic nontransitive forms in Gilbertese or in a related language.

The logic of the first case requires little comment, in that there is no independent motivation in Gilbertese for a phonological process relating *Vi#* and *V#* forms. The second case is perhaps less obvious, but still straightforward: if *C* is thematic, then (given final consonant loss in some ancestor of Gilbertese) there are/were paradigmatically related forms both with and without *C*. The forms without *C* could only have arisen by loss of *C* in some designated position (generally assumed to be word-final position), otherwise *C* would have been lost in all forms. *C* was, therefore, word-final at some stage. Therefore, forms with thematic *C* could only have had a root-final vowel (following the thematic *C*) if some process of final-vowel deletion applied **before** final-consonant deletion. Given that there is no evidence for such a process, one concludes that the *-i* following thematic *C* is a suffix.<sup>18</sup>

It remains to show that there are Gilbertese nontransitive forms in *-i* that meet one or the other of the above conditions. I cannot demonstrate this fact for all the forms in question, since only a few have forms without *-i* or without the thematic consonant in Gilbertese, and I have not been able to find cognates for all the residue. But the evidence available is still substantial. I will review the evidence, first for vowel-final roots, then for thematic roots.

The vowel-final root-class need not delay us long, as I know of only four cases with intransitives without *-i*:

- |      |                                      |   |
|------|--------------------------------------|---|
| (26) | <i>akoa</i> ~ <i>akona</i>           | 'to look after s.th.'                         |
|      | <i>aki</i> [ <i>ako</i> ] <i>ako</i> | 'to be careless' ( <i>aki</i> 'verb negator') |
|      | <i>akoi</i>                          | 'tenderness, to be caring, gentle'            |
|      | <i>inaa</i>                          | 'to adopt/affiance s.o.'                      |
|      | <i>ina</i>                           | 'to have a characteristic mark or sign'       |
|      | <i>inai</i>                          | 'betrothal'                                   |

<sup>18</sup> There is, of course, one other possibility, namely, that *C*-final forms accreted a final vowel at some stage. Such a process has been proposed in some Oceanic languages. If a similar process ever applied in the history of Gilbertese, it was certainly not a general process, since PAn final consonants have been lost in many forms. Given a choice between a sporadic phonetic process sensitive to syllable structure, and an affix that is not fully productive, I prefer the latter.

<i>ioa</i>	'to sift/shake s.th.'
<i>io bon</i>	'to sift dirt'
<i>ioioi</i>	'(act of) sifting'
<i>koa</i>	'to grate s.th.'
<i>koi</i>	'(act of) grating'

For a larger set of -i- transitive verbs, there is evidence that the consonant preceding the -i termination is thematic. The examples in (27) include those for which external cognates demonstrate that the presuffix consonant is thematic, those for which Gilbertese alternations prove it to be thematic, or both.

- (27)
- |                            |                            |
|----------------------------|----------------------------|
| <i>aita</i>                | 'to pull, drag s.th.'      |
| <i>aitiiti</i>             | 'act of pulling, dragging' |
| <i>ai rin</i>              | 'to drag coconut leaves'   |
| <i>aoka</i>                | 'to moisten s.th.'         |
| <i>aoi</i>                 | 'dampness, dew'            |
| <i>ikota</i> <sup>19</sup> | 'to add up, amass s.th.'   |
| <i>ikoti</i>               | 'act of amassing'          |
| <i>iko ben</i>             | 'to collect coconuts'      |
| <i>kinika</i>              | 'to pinch, pluck s.th.'    |
| <i>kiniki</i>              | 'to pinch'                 |
| <i>kiniki uee</i>          | 'to pick flowers'          |
| <i>kin uti</i>             | 'to pick nits'             |
| <i>kinikin</i>             | 'species of stinging ant'  |
- (PMc \*kini[t] 'pinch', Mok *kinij* 'to pinch s.th.')
- |                  |                    |
|------------------|--------------------|
| <i>koroma</i>    | 'to husk'          |
| <i>korom ben</i> | 'to husk coconuts' |
| <i>koro</i>      | 'husking stick'    |
- (PMc \*kosom 'to husk', Mok *kodom* 'to husk s.th.', *kod* 'husking stick')
- |                    |                                    |
|--------------------|------------------------------------|
| <i>motika</i>      | 'to break off, to decide on s.th.' |
| <i>motiki</i>      | 'to decide'                        |
| <i>te tia moti</i> | 'the/a judge'                      |
| <i>onika</i>       | 'to alter, change s.th.'           |
| <i>oniki</i>       | 'to change'                        |
- (Pon *welian* 'to change s.th.', *weliali* 'to change')
- |               |                               |
|---------------|-------------------------------|
| <i>ongira</i> | 'to squeeze, wring out s.th.' |
| <i>ongiri</i> | 'to squeeze'                  |
- (Mok *ungud*, Pon *wengid* 'to squeeze, wring out s.th.', Mok *ongngong*, Pon *wengiwen* 'to squeeze', PMc \*wongis)

<sup>19</sup> The forms *ikota* and *motika* are exceptions to the constraint, noted in §4, barring tautomorphic apical and velar stops in native Gilbertese etyma. Gil *kinika* shows conformity to the constraint.

<i>taima</i>	'to sharpen s.th.'
<i>taim</i>	'to sharpen'
<i>taim biti</i>	'to sharpen knives'
(Mok <i>jaim</i> 'to sharpen s.th.', <i>joaijoai</i> 'to sharpen', Tru <i>seyimi</i> 'to sharpen s.th.', <i>seyim</i> 'to do sharpening, whetstone', PMc * <i>taim</i> , PAn * <i>tazim</i> )	
<i>tangira</i>	'to like, cry for s.th.'
<i>itangitangiri</i>	'to like one another'
<i>tang</i>	'to cry, weep'
(PMc, PAn * <i>tangis</i> )	
<i>tiringa</i>	'to beat, kill, reprimand s.o.'
<i>tiring</i>	'to reprimand'
<i>tiritiri</i>	'brutality'
<i>tauna</i>	'to bury s.th.'
<i>taun</i>	'to bury'
(Mok <i>jaun</i> 'to bury s.th.', <i>joau</i> 'to bury')	
<i>urunga</i>	'to steer (for) s.th.'
<i>urung</i> <sup>20</sup>	'to steer'
(Mok <i>widinge</i> 'to trick s.o.', <i>widing</i> 'to know all the tricks')	

At least three questions remain regarding the *-i* intransitive suffix:

- (i) Is it an innovation of Gilbertese?
- (ii) Why is it restricted to a small set of verbs?
- (iii) What is its origin?

The answer to the first question appears to be negative. There is some evidence of the suffix in other Micronesian languages, at least in the ablaut, reciprocal and reduplicated intransitives of Ponapeic, noted in §2. The ablaut pattern, in which intransitive forms (like Pon *ngked* 'to roof with thatch') have a mid vowel, while the corresponding transitives (like Pon *ngkad* 'to roof s.th. with thatch') have a low vowel suggest a transitive in \*-a and an intransitive in \*-i. I am less familiar with the data from other Micronesian languages, but Tru *seyim* 'to do sharpening, whetstone' seems to be an instance. I must note, however, that it is difficult to assess any argument for high-vowel erasure in other Micronesian languages, given the loss of final vowels.

To the question of the nature of the restrictions on the *-i* intransitive suffix I can offer little insight. I would only observe that in five of the thirteen cases of thematic verbs with intransitive *-i*, the thematic consonant is a nasal. Nasal thematics also seem to be much in evidence in the likely reflexes of \*-i intransitives in other Micronesian languages. If the correlation should prove to be significant, it is still a mystery to me.

To the final and most tantalising question, let me offer the following speculation. The *-i* intransitive termination of Micronesian languages is not a newly discovered suffix, but simply

<sup>20</sup> The semantics of the putative Mokilese cognate implies some metaphoric extension of an original sense closer to that of the Gilbertese. In Harrison (1994) I identified this item, probably incorrectly, as a Polynesian borrowing (cf. Nuclear Polynesian \**uluŋi* 'steer'), having overlooked its Micronesian cognates. Even at the time, I was suspicious of the borrowed status of the item, since Polynesian borrowings into Gilbertese are typically from Samoic languages, and do not usually show loss of postnasal high final vowels.

the *-i* transitive suffix found before object suffixes and, in modern Gilbertese and Marshallese at least, with plural inanimate objects (see Harrison 1978). It is perhaps this last use that provided the bridge for *\*-i* to move into the intransitive domain. In Gilbertese, at any rate, it is only this *-i* form that can appear without any (pronominal or common noun phrase) complement, as in (28):

- (28) *I wareka te boki.*  
 1SG read-3SGNP ART book  
 'I read the/a book.'
- \*I wareka.*
- I warekia.*  
 1SG read-3SG  
 'I read it.'
- I wareki(i) booki.*  
 1SG read-3PL-INAN book  
 'I read the/some books.'
- I wareki(i).*  
 1SG read-3PL-INAN  
 'I read them.'

Furthermore, it is often difficult in Gilbertese to distinguish between noun incorporation nontransitive structures with *-i* intransitive verbs on the one hand, and the corresponding third person plural, inanimate, transitive construction on the other, as in (29):

- (29) *I kiniki uee.*  
 1SG pluck/pluck(-3PL-INAN) flower  
 'I picked (the/some) flowers.'

Only when the common noun complement is bimoric, and thus subject to the process of trimoric contentive lengthening (by which nouns and verbs alone in their phrase are minimally trimoric, and analogous to the 'compensatory lengthening' processes of Ponapean and Trukese) is the contrast clear, as in (30):

- (30) *I korom ben.*  
 1SG husk coconut  
 'I husked coconuts/did coconut-husking.'
- I korom(ii) been.*  
 1SG husk-3PL-INAN coconut  
 'I husked the/some coconuts.'

The *-i* intransitive suffix reflects a change in the nature of transitive marking in Micronesian languages, through which the weight of morphosyntactic transitivity moves from the transitive formant onto the object suffix. Were the change to go to completion, the *-i*-transitive suffix would become part of the verb stem. This change was stillborn in those Micronesian languages that lost final vowels, but remains potentially active in Gilbertese. I have no evidence regarding its likely future progress.

## References

- Burzio, L., 1986, *Italian syntax*. Dordrecht: Reidel.
- Clark, R., 1976, *Aspects of Proto-Polynesian syntax*. Auckland: Linguistic Society of New Zealand.
- Chung, Sandra, 1978, *Case marking and grammatical relations in Polynesian*. Austin: University of Texas Press.
- Durrheimer, Pr R., 1964, *Grammaire Gilbertine*. MS.
- Harrison, S.P., 1978, Transitive marking in Micronesian languages. In S. Wurm and L. Carrington, eds *Second international conference on Austronesian linguistics: Proceedings Fasc 2*, 1067–1127. Canberra: Pacific Linguistics.
- 1982, Proto-Oceanic \*aki(ni) and the Proto-Oceanic periphrastic causative. In S. Wurm and L. Carrington, eds *Second international conference on Austronesian linguistics: Proceedings Fasc 2*, 179–230. Canberra: Pacific Linguistics.
- 1994, Linguistic evidence for Polynesian influence in the Gilbert Islands. In T. Dutton and D. Tryon, eds *Language contact and change in the Austronesian world*, 321–350. Trends in Linguistics: Studies and Monographs, 77. Berlin: Mouton de Gruyter.
- Jackson, Frederick H., 1983, The internal and external relationships of the Trukic languages of Micronesia. PhD dissertation, University of Hawai'i.
- Levin, B., 1993, *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press.
- Pawley, Andrew K., 1973, Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103–188.
- Perlmutter, David and Paul Postal, 1984, The I-advancement exclusiveness law. In David Perlmutter and Carol Rosen, eds *Studies in relational grammar 2*:81–125. Chicago: University of Chicago Press.
- Pustejovsky, J., 1995, *The generative lexicon*. Cambridge: MIT Press.
- Rehg, Kenneth L., 1981, *Ponapean reference grammar*. Honolulu: University of Hawai'i Press.

# *On the morphological status of thematic consonants in two Oceanic languages*

---

FRANTISEK LICHTENBERK

Many Oceanic languages contain a derivational pattern relating transitive verbs and the bases from which they are derived such that the derived form contains a suffix not found in the base, and commonly there is a consonant found in the derived form that is absent from the base. This study reviews the status of these ‘thematic’ consonants in two representative languages and evaluates the two opposing solutions proposed in the literature to date: the conjugation analysis, whereby the consonants are considered part of a suffix with lexically determined allomorphs; and the phonological analysis, whereby the consonants are considered part of the stem that are deleted under general phonological conditions. Careful analysis of data from Toqabaqita and Manam suggest instead a morpholexic analysis, where the thematic consonants are analysable as morphs only on formal grounds, not on semantic grounds; they are not signs. It is only the combination of a base, a thematic consonant, and/or a transitive suffix that is meaningful. Lexemes, as signs, may be combinations of signs and nonsigns, something explicitly recognised by a lexeme-based view of morphology (Aronoff 1994). To the extent that there are some associations between thematic consonants and semantic fields, it looks as though the thematic consonants were morphs in search of meanings.

## 1 Introduction<sup>1</sup>

Many Oceanic languages contain a derivational pattern relating transitive verbs and the bases from which they are derived such that the derived form contains a suffix not found in the base, and commonly there is a consonant found in the derived form that is absent from the source base. Examples (1) and (2) illustrate, with the relevant consonants in boldface.

---

<sup>1</sup> It gives me great pleasure to dedicate this paper to Byron Bender. Byron was the Chair of the Department of Linguistics at the University of Hawai‘i during my PhD studies there, and I always appreciated his support. In 1994 I delivered a somewhat different version of this paper at the Seventh International Conference on Austronesian Linguistics in Leiden, and Byron provided me with valuable comments. I am also grateful to Laurie Bauer and Wayne Lawrence for their comments on that version of the paper.

Following the tradition in Oceanic linguistics, I will use the term **thematic consonants** to refer to consonants that occur in transitive verbs, for example the *f* in (1b) and the *r* in (2b):<sup>2</sup>

## TOQABAQITA

- (1) a. *kuqu* (intr) 'drink'  
 b. *kuqufia* (tr) 'drink'

## MANAM

- (2) a. *mongo* (intr) 'spit'  
 b. *mongoraqi* (tr) 'spit out'

The purpose of this paper is to investigate the morphological status of the thematic consonants in two Oceanic languages, Toqabaqita and Manam. Are the consonants part of the suffix, part of the base, or neither? That is, should the transitive verbs in (1b) and (2b) respectively be analysed as *kuqu-fī-a* and *mongo-raq-i*, or as *kuquf-i-a* and *mongor-aq-i*, or in some other way?

The problem of the morphological status of the thematic consonants in these two languages is reminiscent of the well-known Maori case, extensively discussed in the literature (for example Hale 1973, 1991; McCarthy 1981; Sanders 1990, 1991; Blevins 1994). In Maori, the passive and nominalised forms of verbs contain a variable (thematic) consonant not found in the corresponding active forms. The examples in (3) demonstrate a few passive forms:

- |     |                |                   |          |
|-----|----------------|-------------------|----------|
| (3) | ACTIVE         | PASSIVE           |          |
|     | <i>koohuru</i> | <i>koohurutia</i> | 'murder' |
|     | <i>huri</i>    | <i>hurihia</i>    | 'turn'   |
|     | <i>tanu</i>    | <i>tanumia</i>    | 'bury'   |
|     | <i>mau</i>     | <i>mauria</i>     | 'carry'  |
- (Hohepa 1967:59, 106–110)

Which of the thematic consonants a given verb will take is not predictable on phonological (or any other) grounds, although Blevins (1994) has argued that the passive suffix has two default forms whose choice is determined by the number of morae of the base. With some verbs, the default variant of the suffix is replaced by another allomorph. In a pioneering discussion of the Maori case, Hale (1973) presented two kinds of analysis, and the subsequent attempts to deal with the data have usually argued the pros and the cons of the two basic types of approach. In the so-called 'conjugation analysis', the thematic consonants are considered part of the passive and the nominalising suffixes, and both suffixes have a number of allomorphs whose distribution is determined lexically. In the other, 'phonological', analysis the thematic consonants are part of the stem, and they are deleted by a general phonological rule when no suffix follows: there are no closed syllables in Maori. Most analysts, on a

<sup>2</sup> The Toqabaqita and Manam data come from my field notes. Orthographic *q* represents /ŋ/, *ng* represents /ŋ/, and *th* represents /θ/. The Fijian forms in §7 are given in Fijian orthography. Toqabaqita transitive verbs take object suffixes under certain conditions. Here transitive verbs will be given with the third person singular object suffix *-a*, although for the sake of brevity they will usually be glossed without reference to objects (thus 'drink' (tr) rather than 'drink it'). And, also for brevity, although Manam transitive verbs always contain an object suffix, they will usually be glossed without reference to objects ('spit out' rather than 'spit it out'). The following abbreviations are used in the glosses: 1 – first person, 3 – third person, CAUS – causative prefix, intr – intransitive (verb), NOM – nominaliser, O – object suffix, RED – reduplication, REP – repetition, SG – singular, s.o. – someone, s.th. – something, TC – thematic consonant, tr – transitive (verb), TR – transitive suffix.



variety of grounds, favour the conjugation analysis, among them Hohepa (1967), Hale (1973, 1991), (Kaye 1975), (McCarthy 1981), and also Blevins (1994). The phonological analysis has been argued for by Sanders (1990, 1991).

It is not my intention here to enter the debate concerning the Maori case, and I will leave it as an empirical question whether the analysis of the thematic consonants to be presented here for Toqabaqita and Manam can be extended to Maori and other languages. For convenience, I will adopt the terms 'conjugation analysis' and 'phonological analysis' in the way they have been used in the discussion of Maori. Under the phonological analysis, the thematic consonants are considered to be part of the base, as in Toqabaqita *kuquf* 'drink' in (1) above and Manam *mongor* in (2). To get the transitive forms, the suffixes *-i-a* and *-aq-i* respectively are added. To get the intransitive forms, the thematic consonants are deleted. On the conjugation analysis, the thematic consonants are considered part of the affixes, thus Toqabaqita *kuqu-fi-a* and Manam *mongo-raq-i*. Since, as will be seen later, both languages have a number of thematic consonants, there would be a number of variants of the suffixes (but there is also some additional variation that will be mentioned later).

I will present evidence that neither the phonological nor the conjugation analysis is appropriate for the two languages. Instead, I will argue for what I will call a **morpholexic** analysis. The first part of the term serves to highlight the aspect of the analysis according to which the thematic consonants are separate morphs. The second part highlights the lexemic framework adopted here, in which it is the lexeme rather than the morpheme that is the minimal sign. Although the thematic consonants are morphs, they do not carry meanings: they are empty morphs (but that is not to deny some tangential semantic links, as will be discussed later).

The paper is structured as follows. In §2, I present some background information that will be relevant to subsequent discussion. In §3, variation in thematic consonants is discussed. In §4, evidence is given against the phonological analysis for both languages, and in §5 evidence is given against the conjugation analysis, also for both languages. Section 6 provides arguments for the morpholexic analysis; §7 discusses associations between thematic consonants and verb meanings, and §8 contains concluding remarks.

Unless I am specifically discussing the phonological analysis, in which case the thematic consonants will be written as part of the base, or unless I am specifically discussing the conjugation analysis, in which case the thematic consonants will be written as part of the transitive or object suffixes, I will write the thematic consonants separately. This is partly to steer a neutral course between the phonological and the conjugation analyses and partly in anticipation of some aspects of the analysis to be adopted. I use the term **base** rather than **root** because the form that serves as the input into the transitive derivation may itself be complex, such as a compound verb. On occasion, I will use the term **inner base**, which is the base to which the thematic consonant is added. This new base then takes the transitive and/or the object suffixes.

Both the phonological and the conjugation types of analysis suffer from serious shortcomings; in some cases evidence against one type of analysis appears to be evidence for the other type. There is some overlap between the two languages with respect to the kinds of evidence against the two types of analysis, but there are also differences. Nevertheless, there is convincing evidence for rejecting both types of analysis for both languages.

## 2 Diachronic and synchronic background

Pawley (1973) has reconstructed two transitive suffixes for Proto Oceanic, \*-i and \*-aki(ni). These are sometimes referred to as **short** and **long transitive suffixes**, respectively. Both suffixes survive—mostly with some kind of metamorphosis—in Toqabaqita and Manam. In Toqabaqita, the short suffix survives intact, while the long suffix is reflected in two variants: *-aqi* and *-ani*. The latter is used when an object-indexing suffix follows, and the former is used when no object suffix follows. With one type of verb, the one relevant in the context of this paper, object suffixes exist only for the third person (singular, dual and plural). Examples (4) and (5) show the variation in the form of the long suffix depending on the presence or absence of an object-indexing suffix:

- (4) *fuqaro-ng-ani-a*  
whisper-TC-TR-3SG:O  
'whisper about him/her/it'
- (5) *fuqaro-ng-aqi nau*  
whisper-TC-TR 1SG  
'whisper about me'

There is no variation in the form of the short transitive suffix.

- (6) *karu-m-i-a*  
scratch-TC-TR-3SG:O  
'scratch him/her/it'
- (7) *karu-m-i nau*  
scratch-TC-TR 1SG  
'scratch me'

In Manam, both the former short transitive suffix \*-i and the *i* of the erstwhile long transitive suffix \*-aki have been reanalysed as the third person singular object suffix; see (8) and (9) respectively:

- (8) *ado-r-i*  
be.straight-TC-3SG:O  
'straighten (tr)'
- (9) *qawa-t-aq-i*  
know-TC-TR-3SG:O  
'know (about) (tr)'

Unlike Toqabaqita, Manam has object-indexing suffixes for all persons and numbers; this will be discussed in detail in §5.5.

For convenience, I will use the terms **short transitive verbs** and **long transitive verbs** to refer to verbs that carry a reflex of the erstwhile short or long transitive suffix, respectively (even though in Manam the reflex of the short transitive suffix does not function as a transitive suffix).

In both languages, in addition to the various suffixes, the derivation of a transitive verb may also involve a causative prefix. This is illustrated in (10) for Manam:

- (10) *aqā-salaga-t-i*  
 CAUS-be.long-TC-3SG:O  
 'make long, longer'

A transitive verb need not contain a thematic consonant. This is true both of short transitives and long transitives in both languages:

TOQABAQITA

- (11) *tala-i-a*  
 road-TR-3SG:O  
 'lead (s.o.)'

MANAM

- (12) *laqu-aq-i*  
 go-TR-3SG:O  
 'take, carry to a place away from deictic centre'

(Without the transitive suffix, the form of the verb is *laqo*. Various kinds of vowel alternation are found in some other verbs.)

Both in Toqabaqita and in Manam, some bases may be used to form both short and long transitives. In some cases, the thematic consonant is the same in the short and the long transitive verbs; in others, the consonants are different (or one form has a thematic consonant and the other does not); see (13) and (14) respectively:

MANAM

- (13) a. *mongo-r-i*  
 spit-TC-3SG:O  
 'spit at'
- b. *mongo-r-aq-i*  
 spit-TC-TR-3SG:O  
 'spit out'

TOQABAQITA

- (14) a. *fita-l-i-a*  
 run-TC-TR-3SG:O  
 'run around in (rain)' (little children's favourite activity)
- b. *fita-t-ani-a*  
 run-TC-TR-3SG:O  
 'take s.o. (somewhere) in a hurry' (e.g. an injured person to the clinic)

As variation in thematic consonants will be used as evidence against both the phonological and the conjugation analyses, a detailed discussion is necessary.

### 3 Variation in thematic consonants

In both Manam and Toqabaqita, most bases are associated with unique thematic consonants.<sup>3</sup> In many cases, only one transitive form—short or long—exists. And in those cases where both a short and a long transitive verb are derived from the same base, the thematic consonants are often the same:

#### MANAM

- (15) a. *qulena-r-i*  
vomit-TC-3SG:O  
'vomit on'
- b. *qulena-r-aq-i*  
vomit -TC-TR-3SG:O  
'vomit out'

#### TOQABAQITA

- (16) a. *teo-f-i-a*  
lie-TC-TR-3SG:O  
'lie on top of'
- b. *teo-f-ani-a*  
lie-TC-TR-3SG:O  
'lay down'

However, in both languages there are also instances where one and the same base is associated with more than one thematic consonant, or there is variation between the presence and absence of a thematic consonant. In Manam, it is the latter kind of variation that is the typical case. It may be either the short or the long transitive form that has no thematic consonant, while the other form does have one:

- (17) a. *suburawa-i*  
sweat-3SG:O  
'pass one's sweat on; soil with sweat'
- b. *suburawa-ng-aq-i*  
sweat-TC-TR-3SG:O  
'sweat out'
- (18) a. *bole-s-i*  
tell.lie-TC-3SG:O  
'tell lies to'
- b. *boli-aq-i*  
tell.lie-TR-3SG:O  
'tell lies about'

(The short transitive form *bolesi* has *e* rather than *i*.)

<sup>3</sup> Here and elsewhere I speak of verbs being **associated** with thematic consonants regardless of whether the thematic consonants are to be considered part of the inner base or not. In a similar vein, I will speak of thematic consonants being associated with suffixes.

There is at least one case where an intransitive verb is associated with two thematic consonants in Manam. The verb *pura* 'arrive' has a short transitive counterpart with *q* as the thematic consonant: *pura-q-i* 'deliver'. *Pura* is also compounded with *laqo* 'go': *laqo-pura* 'arrive at a place away from deictic centre'; and this compound has a long transitive counterpart with *ng* as the thematic consonant: *laqo-pura-ng-aq-i* 'deliver to a place away from deictic centre'. The verb *laqo* 'go' itself has a long transitive variant with no thematic consonant (and *u* in place of *o*): *laqu-aq-i* 'take, carry to a place away from deictic centre'.

In Toqabaqita, the extent of variation in thematic consonants is greater than in Manam, and several different types of variation can be distinguished. In one, there is a regular correspondence between a thematic consonant in a short transitive and another thematic consonant in the corresponding long transitive which is the consequence of an earlier sound change. In the history of Toqabaqita, a sound change took place whereby \**s* became *t* before nonhigh vowels and continued as *s* before high vowels (Pawley 1972, Lichtenberk 1988). This split has resulted in an alternation between *s* as a thematic consonant in short transitives (before *i*) and *t* as its counterpart in long transitives (before *a*), for example:

- (19) a. *moa-s-i-a*  
vomit-TC-TR-3SG:O  
'vomit on'
- b. *moa-t-ani-a*  
vomit-TC-TR-3SG:O  
'vomit out'

Although, owing to developments postdating the split of \**s*, *s* and *t* are separate phonemes in Toqabaqita today, I am not aware of any short transitives with *t* as the thematic consonant or of any long transitives with *s* as the thematic consonant.

In some other cases, there is variation in thematic consonants between the short and the long transitive forms without a regular correspondence:

- (20) a. *liu-f-i-a*  
walk/roam-TC-TR-3SG:O  
'walk, roam around (a place)'
- b. *liu-t-an-i-a*  
walk/roam-TC-TR-3SG:O  
'come and destroy, plunder (food in s.o.'s garden; out of spite, as punishment), and walk away'

(There is also a long transitive form *liu-f-ani-a*; see (27b) further below.)

- (21) a. *dola-l-i-a*  
disappear-TC-TR-3SG:O  
'mix (s.th.) in, dissolve (s.th.) in' (the stuff into which s.th. is mixed or in which s.th. is dissolved as direct object)
- b. *dola-ng-ani-a*  
disappear-TC-TR-3SG:O  
'lose'

I am not aware of variation between the presence and absence of a thematic consonant between short and long transitives in Toqabaqita; however, there is a different kind of variation between the presence and absence of a thematic consonant. Besides short transitive

verbs, which contain a reflex of the Proto Oceanic transitive suffix \*-i, and long transitive verbs, which contain a reflex of the Proto Oceanic transitive suffix \*-aki(ni), Toqabaqita has a third type of transitive verb that contains neither the reflex of \*-i nor the reflex of \*-aki(ni). The object suffixes are added directly to the source base; and there is no thematic consonant. I will refer to such transitive verbs as **bare**. For example:

- (22) *duqu-a*  
ridgepole-3SG:O  
'install ridgepole in (a house)'

A bare transitive verb may have a short or a long transitive counterpart, or even both:

- (23) a. *fono-a*  
be.shut-3SG:O  
'shut'
- b. *fono-s-i-a*  
shut-TC-TR-3SG:O  
'install (the last thatching panel, when roofing a house)'
- (24) a. *lole-a*  
tell.lie-3SG:O  
'lie to'
- b. *lole-t-ani-a*  
tell.lie-TC-TR-3SG:O  
'lead, entice s.o. to a place under a pretext in order to kill them'

As shown in (21) above, the intransitive verb *dola* 'disappear, be lost' has a short transitive variant with *l*: *dola-l-i-a* 'mix (s.th.) in, dissolve (s.th.) in' (the stuff into which s.th. is mixed or in which s.th. is dissolved as direct object, and the person doing the mixing as subject); and a long transitive variant: *dola-ng-ani-a* 'lose'; and there is also a bare transitive variant: *dola-a* 'dissolve, mix (itself, oneself) in' (the stuff, group into which s.th. or s.o. is dissolved, mixed as direct object, and the stuff or person that is dissolved or mixed as subject).

Lastly, in Toqabaqita there is also variation in thematic consonants between two short transitives or between two long transitives; that is, there is more than one short transitive form or more than one long transitive form derived from the same base distinguished only by the thematic consonants. In some cases, the variants are synonymous:

- (25) a. *fuqaro-ng-ani-a*  
whisper-TC-TR-3SG:O  
'whisper about'
- b. *fuqaro-t-ani-a*  
whisper-TC-TR-3SG:O  
'whisper about'

In others, there is a difference in meaning:

- (26) a. *ngata-l-i-a*  
speak-TC-TR-3SG:O  
'speak about, discuss'

- b. *ngata-f-i-a*  
 speak-TC-TR-3SG:O  
 'speak angrily to, berate'
- (27) a. *liu-t-ani-a*  
 walk/roam-TC-TR-3SG:O  
 'come and destroy, plunder (food in s.o.'s garden; out of spite, as punishment),  
 and walk away'
- b. *liu-f-ani-a*  
 walk/roam-TC-TR-3SG:O  
 'befall, "visit" (of a disease or death befalling a person as retribution for  
 taboo behaviour)'

There is also a short transitive verb *liu-f-i-a* 'walk, roam around (a place)' (see 20a) and even a bare transitive verb *liu-a* 'pass the time for (doing s.th.), be too late for (doing s.th.)' (the object is a clause encoding the situation whose time has passed, and there is no lexical subject).

The existence of variation in thematic consonants will be used as evidence against both the phonological and conjugation analyses.

#### 4 Evidence against the phonological analysis

On the phonological analysis as originally formulated by Hale (1973), the thematic consonant would be part of the inner base, and it would be deleted when no suffix follows. Toqabaqita does not permit word-final consonants at the underlying level,<sup>4</sup> and Manam permits only the nasals /m/ and /ŋ/ word finally (the two being in free variation). There is strong evidence against the phonological analysis in both languages.

##### 4.1 Variation between the presence and absence of a thematic consonant

Both languages contain instances of more than one transitive verb derived from the same base, one of which contains no thematic consonant, while another form does contain a thematic consonant. There are many cases of this sort in Toqabaqita that involve bare verbs, namely, verbs that contain neither a transitive suffix nor a thematic consonant. The form in (28a) is a bare verb, while the form in (28b) is a short transitive verb derived from the same base.

- (28) a. *kwethu-a*  
 be.alight/burn-3SG:O  
 'light (a torch)'
- b. *kwethu-f-i-a*  
 be.alight/burn-TC-TR-3SG:O  
 'of a fire: consume by burning'

<sup>4</sup> Closed syllables do optionally arise at the phonetic level, but such are not relevant to the present discussion.

If, assuming the phonological analysis, the stem were *kwethuf*, the final *f* would be deleted in order to form the intransitive verb *kwethu*. However, the consonant would also need to be deleted in the formation of the bare transitive *kwethu-a* even though the *f* would not be word final. There would be no phonological motivation for this deletion. Similar variation, this time between a bare transitive and a long transitive, is found in *lole-a* 'lie to' and *lole-t-ani-a* 'lead, entice s.o. to a place under a pretext in order to kill them' (*lole* 'tell a lie'). Variation between no thematic consonant in a bare transitive and a thematic consonant in a short or a long transitive is not at all rare.

In Manam there are cases of variation between no thematic consonant in a short transitive verb and a thematic consonant in a long transitive verb, or between a thematic consonant in a short transitive and no thematic consonant in a long transitive. The forms in (29) illustrate the former kind of variation:

- (29) a. *suburawa-i*  
sweat-3SG:O  
'pass one's sweat on; soil with sweat'
- b. *suburawa-ng-aq-i*  
sweat-TC-TR-3SG:O  
'sweat out'

In order to derive the intransitive verb *suburawa* 'sweat', the final *ng* of *suburawang* would be deleted. However, there would be no phonological motivation for deleting the *ng* in the short transitive form.

In the case of *suburawang* (and other verbs that are the same in this respect), there is an additional problem. While Manam does not permit most consonants to occur word-finally, two of the three nasals, /m/ and /ŋ/ (but not /n/), do occur word-finally, as in *Manam* 'Manam', *mang* 'bird' and *malabong* 'flying fox'. There is then no phonological reason why the *ng* could not be present in the intransitive verb, but it is not. The form is *suburawa*, not \**suburawang*.

The forms in (30), without a thematic consonant in the noun (a) and with a thematic consonant in the related transitive verb (b) illustrate the inadequacy of the phonological analysis in a different way:

- (30) a. *tina*  
'mother'
- b. *tina-m-i*  
mother-TC-3SG:O  
'regard, treat s.o. as one's own mother'

Like other kinship terms, the term for 'mother' obligatorily carries a possessive suffix indexing the 'possessor' (*tina-gu* 'my mother', *tina-ng* 'your (SG) mother' and so on), except for third person singular possessors, which are not overtly indexed: *tina* 'his/her mother'. On the phonological analysis, the final *m* of the base *tinam* would be absent from the noun; but there is no phonological motivation for this. First, *m* can occur word-finally, and so the form *tinam* 'his/her mother' would be phonologically acceptable. Second, Manam does permit homorganic NC clusters across syllable boundaries; as in /*lunguma*/ 'person from a village other than one's own'. That is, the form /*tinangu*/ 'my mother' (with nasal assimilation) would be phonologically acceptable. Finally, although consonant clusters other than heterosyllabic homorganic NC sequences are not permitted, Manam has a rule of epenthesis



whereby *i* is used to break up impermissible clusters that arise through concatenation of morphemes; as in *mata-ng* ‘your (SG) eye(s)’, but *mata-dang-ing* ‘your (SG) tears’ (lit. ‘your eye-water’). There is then no phonological reason why a form like *tinam-ing* ‘your (SG) mother’ could not exist.

#### 4.2 Variation in thematic consonants

As pointed out in §3, in both languages there are bases associated with more than one thematic consonant. This kind of variation is more common in Toqabaqita than in Manam. Three basic types of variation may be distinguished: (i) one thematic consonant in a short transitive and another thematic consonant in a long transitive; (ii) two (or more) thematic consonants in short or in long transitives where the verbs are not synonymous; and (iii) two (or more) thematic consonants in short or in long transitives where the variants are synonymous. I will discuss the three types of variation in turn, starting with (iii).

The forms in (31) from Toqabaqita illustrate the existence of two synonymous long transitive forms distinguished only by the thematic consonants:

- (31) a. *fuqaro-ng-ani-a*  
whisper-TC-TR-3SG:O  
‘whisper about’
- b. *fuqaro-t-ani-a*  
whisper-TC-TR-3SG:O  
‘whisper about’

Cases like these do not, of themselves, pose a serious problem for the phonological analysis. One could assume that the morpheme ‘whisper’ has two allomorphs, *fuqarong* and *fuqarot*.

The situation is quite different with cases of type (ii): two (or more) short transitives or two (or more) long transitives that are distinguished by the thematic consonants and that are semantically related to each other but cannot be said to be synonymous. This is illustrated in (32) from Toqabaqita:

- (32) a. *ngata-l-i-a*  
speak-TC-TR-3SG:O  
‘speak about, discuss’
- b. *ngata-f-i-a*  
speak-TC-TR-3SG:O  
‘speak angrily to, berate’

Would the form of the morpheme ‘speak’ be *ngatal* or *ngataf*? Alternatively, one could say that there are in fact two morphemes involved, *ngatal* ‘speak’ and *ngataf* ‘speak angrily’. Given the fact that *ngata* means ‘speak’ without any necessary implication of ‘angrily’, one could argue that *ngata* is derived from *ngatal* rather than from *ngataf*, and that the existence of the two nearly homophonous and semantically closely related forms *ngatal* and *ngataf* is no more than an accident. The problem is that this is not an isolated case. The forms in (33) illustrate this with long transitives:

- (33) a. *liu-t-ani-a*  
walk/roam-TC-TR-3SG:O  
'come and destroy, plunder (food in s.o.'s garden; out of spite, as punishment),  
and walk away'
- b. *liu-f-ani-a*  
walk/roam-TC-TR-3SG:O  
'befall, "visit" (of a disease or death befalling a person as retribution for  
taboo behaviour)'

One would have to claim that, of the two morphemes *liut* and *liuf*, only one is related to the intransitive verb *liu* 'walk, roam' and that only one—and not necessarily the same one—is related to the bare transitive *liu-a* 'pass the time for (doing s.th.), be too late for (doing s.th.)' (see (27) above and the paragraph following it). This kind of analysis would result in a number of nearly homophonous pairs of morphemes that exhibit various kinds of semantic relatedness.

A somewhat similar problem would arise with cases of type (i), those where there is one thematic consonant in a short transitive form and a different thematic consonant in the corresponding long transitive form. This type of situation also is not at all unusual in Toqabaqita.

- (34) a. *lofo-q-i-a*  
fly/jump-TC-TR-3SG:O  
'pounce on, jump in order to catch' (e.g. a cat pouncing on a mouse)
- b. *lofo-t-ani-a*  
fly/jump-TC-TR-3SG:O  
'of an adult bird: assist (a fledgling) in flying'

One alternative would be to say that the morpheme 'fly, jump' has two allomorphs, *lofoq* and *lofot*, whose distribution is conditioned morphologically: *lofoq* before the short transitive suffix, and *lofot* before the long transitive suffixes *-ani* and *-aqi*. Or, as with cases of type (ii), one could posit two mutually independent forms *lofoq* and *lofot*, only one of which would be related to the intransitive verb *lofo*. Which one? The decision would appear to be arbitrary.

A somewhat different situation exists in the case of the alternation between *s* in short transitives and *t* in long transitives, which is the result of an earlier sound change (see §3); for example:

- (35) a. *feqe-s-i-a*  
defecate-TC-TR-3SG:O  
'defecate on'
- b. *feqe-t-ani-a*  
defecate-TC-TR-3SG:O  
'defecate out, pass (faeces), pass with faeces'  
(The intransitive verb 'defecate' is *feqa*.)

Although the *s-t* alternation is fully regular in the case of thematic consonants, elsewhere *s* does occur before *a*, and *t* does occur before *i*; as in *sadi* 'be hard' and *titiu* 'be bent down'. Either one would have to say that *feqes* and *feqet* are coallomorphs whose distribution has to be stipulated in morphological terms, or one would have to claim that these are two separate morphemes that just happen to be very similar phonologically and closely related

semantically. However, in the latter case the fact that there is regular alternation between *s* and *t* would be just another accident.

There are even cases of more than two thematic consonants being associated with a given base: *oli* (intr) ‘return’, *oli-s-i-a* ‘answer s.o.’ and *oli-t-ani-a* or *oli-f-ani-a* ‘return s.th., take back’. The problems with the phonological analysis are even more extensive here.

Variation in thematic consonants is less common in Manam, but it does occur. The forms in (36) show alternation between *t* and *ng* in two long transitives. In the latter case, the verb also contains the causative prefix *aq-*:

- (36) a. *lewa-t-aq-i*  
 be.lost/disappear-TC-TR-3SG:O  
 ‘forget’
- b. *aq-lewa-ng-aq-i*  
 CAUS-be.lost/disappear-TC-TR-3SG:O  
 ‘cause to disappear’

Either one would have to claim that the morpheme ‘be lost, disappear’ has two allomorphs, one of which takes the causative prefix, or one would have to posit two mutually independent morphemes that are related semantically and just happen to be near homophones.

Clearly, the phonological analysis has little to recommend it. It would necessitate positing sets of morphemes that are semantically related and are near homophones, and/or it would require some morphological conditioning, which would defeat its purpose.

### 4.3 Reduplication

Both Toqabaqita and Manam have more than one reduplication process. The ones that are relevant here involve suffixation: part of the base is reduplicated and suffixed to the original base. In Toqabaqita it is the last syllable that is reduplicated; in Manam, the last two syllables. With verbs, reduplication serves to express progressive, habitual and persistent aspects.

#### TOQABAQITA

- (37) *angi-ngi*  
 cry-RED  
 ‘keep on crying’

#### MANAM

- (38) *malipi-lipi*  
 work-RED  
 ‘keep on working’

When a transitive verb associated with a thematic consonant is reduplicated, the thematic consonant is not reduplicated; it follows the reduplication.

#### TOQABAQITA

- (39) *angi-ngi-s-i-a*  
 cry-RED-TC-TR-3SG:O  
 ‘keep on crying for s.o., s.th.’

## MANAM

- (40) *malipi-lipi-t-aq-i*  
 work-RED-TC-TR-3SG:O  
 'keep on working on s.th.'

If the thematic consonant were part of the inner base, one would expect it to be included in the reduplication; however, it is not. This suggests that it is not part of the inner base.

Wayne Lawrence has suggested (pers. comm., 1994) that perhaps the reason why the thematic consonant is not reduplicated in forms such as those in (39) and (40) is that impermissible consonant clusters would result: Toqabaqita \*angisngisia and Manam \*malipitlipitaqi. Toqabaqita permits no consonant clusters at the underlying level, and Manam permits only heterosyllabic homorganic NC clusters. While this argument has some merit in the case of Toqabaqita, it is not quite applicable in the case of Manam. As pointed out in §4.1, Manam has a rule of *i*-epenthesis that breaks up impermissible consonant clusters that arise through concatenation of morphemes. Given the existence of *i*-epenthesis, one might expect \*malipitlipitaqi, but this is wrong.

There is further evidence against the phonological analysis from Manam reduplication. Since Manam does permit heterosyllabic NC clusters, there is no reason why a nasal thematic consonant could not be reduplicated if followed by a single consonant; but this does not happen. This is illustrated in (41):

- (41) *rapu-rapu-ng-i*  
 wait-RED-TC-3SG:O  
 'keep on waiting for'

If the thematic consonant were part of the inner base, one would expect \*rapunrapungi (with assimilation of the first nasal), but although there is no phonological prohibition against this form (note *qunr-aq-i* 'bury'), it is wrong.

In Toqabaqita, evidence for or against the phonological analysis would come from verbs that consist exclusively of vowels followed by a thematic consonant, such as the hypothetical form \*\*ao-f-i. If the reduplicated form were to be \*\*aofaofi, this would support the phonological analysis; however, if the form were to be \*\*aoaofi, this would count as evidence against the phonological analysis. Unfortunately, I have come across no verbs of the requisite shape.

#### 4.4 Phonotactics

If the thematic consonants were to be considered part of the inner base, there would be an odd skewing in the phonotactic patterns in both languages. In Manam, verbs with thematic consonants would be the only lexical morphemes ending in an obstruent; all other lexical morphemes would end either in a vowel or in *m* or *ng*. In Toqabaqita, verbs with a thematic consonant would be the only lexical morphemes whose basic form would end in a consonant. The basic forms of all other lexical morphemes would end in a vowel. (Certain nouns undergo metathesis when suffixed with one of the third person singular possessive suffixes, which results in a morpheme form with a final consonant; but elsewhere the morpheme ends in a vowel, thus *suli-ku* 'my back' and *suil-a* 'his/her back'.)

#### 4.5 Summary of the evidence

There are various kinds of evidence against the phonological analysis: absence of thematic consonants where there is no phonological reason for their absence, alternation in thematic consonants, reduplication and phonotactics. Although they do not all carry equal weight—the first two are stronger than the other two—together they present convincing evidence for rejecting the phonological analysis.

### 5 Evidence against the conjugation analysis

#### 5.1 Introduction

On the conjugation analysis, the thematic consonants are part of the suffixes: in Toqabaqita they would be part of the two transitive suffixes, and in Manam they would be part of the object suffixes (see §5.5 below for more detail) and the transitive suffix. Each suffix would have as many allomorphs as there are thematic consonants, and in some cases an additional allomorph without a thematic consonant. In neither language would the distribution of the allomorphs be specifiable in phonological terms.

First, there are homophonous or near homophonous bases that are associated with different thematic consonants:

#### TOQABAQITA

- (42) *kwala-ng-i*  
swear-TC-3SG:O  
'swear at'

but:

- (43) *kwala-f-i*  
have.offspring-TC-3SG:O  
'give birth to'

#### MANAM

- (44) *mongo-r-i*  
spit-TC-3SG:O  
'spit at'

but:

- (45) *rongo-t-i*  
have.disagreeable (acid/bitter) taste-TC-3SG:O  
'e.g. of food or a stinging plant: cause an itching, burning sensation in s.o.'

Second, one and the same base may be associated with a thematic consonant and no consonant, or with more than one thematic consonant, as demonstrated in §4.2 and §4.3 respectively.

There are various kinds of evidence against the conjugation analysis.

## 5.2 Segmentability

There would be recurrent partial similarities across the allomorphs of each suffix as well as between the two types of suffix:

- (46) Toqabaqita: *-fi, -mi, -ngi, -si, -ni -li, -ri, -qi, -i*  
                   *-faqi, -maqi, -ngaqi, -taqi*  
 Manam: *-ti, -si, -ri, -mi, -ni, -ngi, -qi, -li, -i*  
                   *-taq-i, -saq-i, -raq-i, -maq-i, -naq-i, -ngaq-i, -qaq-i, -aq-i*

Because of the recurrent partial similarities, the thematic consonants are segmentable from the following material and identifiable as separate morphs. (Note that one could also argue that the *i* in the longer forms in Toqabaqita also is to be segmented off for the same reason. I will return to this at the end of §6.)

## 5.3 The sets of thematic consonants

In neither language is the relation between the set of the thematic consonants that are associated with the shorter suffixes and the set associated with the longer suffixes random. In Manam, the set of the thematic consonants associated with the longer suffixes is a proper subset of the set associated with the shorter suffixes. With one exception, this is also true of Toqabaqita, the exception being *t* associated with the longer suffix but not with the shorter suffix. However, as pointed out in §3, this is a consequence of an earlier sound change: *s* ceased to occur before *a*, where *t* occurred instead. We can say then that in both languages the set of the thematic consonants associated with the long suffixes is **circumscribed** by the set associated with the short suffixes: only consonants that occur in the latter set occur in the former set (though not all of them do), or, in one case, there is a regular correspondence between a consonant in the former set and a consonant in the latter set.

If the thematic consonants were part of the suffixes, this kind of relation between the two sets of thematic consonants would be no more than an accident. (Under the phonological analysis this issue would not arise.) In neither language does the complete set of thematic consonants correspond to the total inventory of consonants. As Tables 1 and 2 show, in both languages the sets of thematic consonants are smaller than the total inventories of consonants: in Toqabaqita nine out of seventeen, and in Manam eight out of thirteen. That is, in both languages there is room for the longer suffixes to be associated with thematic consonants not associated with the shorter suffixes; however, this does not happen (with the exception of the *s*–*t* correspondence in Toqabaqita). The two sets of thematic consonants are not completely independent of each other, something that the conjugation analysis would be silent about.

**Table 1:** Consonant inventory and thematic consonants in Toqabaqita

	<i>t</i>	<i>k</i>	<i>kw</i>	<i>q</i>	<i>b</i>	<i>d</i>	<i>g</i>	<i>gw</i>	<i>f</i>	<i>th</i>	<i>s</i>	<i>m</i>	<i>n</i>	<i>ng</i>	<i>l</i>	<i>r</i>	<i>w</i>
TC (short suffix)				+					+		+	+	+	+	+	+	
TC (long suffix)	+								+			+		+			

**Table 2:** Consonant inventory and thematic consonants in Manam

	<i>p</i>	<i>t</i>	<i>q</i>	<i>b</i>	<i>d</i>	<i>g</i>	<i>m</i>	<i>n</i>	<i>ng</i>	<i>s</i>	<i>z</i>	<i>l</i>	<i>r</i>
TC (short suffix)		+	+				+	+	+	+		+	+
TC (long suffix)		+	+				+	+	+	+			+

#### 5.4 Alternation in thematic consonants coupled with semantic differences

As exemplified in §4.2 and §4.3, both languages exhibit alternations between presence and absence of a thematic consonant and alternations between two, or even more, thematic consonants. In some cases, the forms are synonymous, as in Toqabaqita *fuqarongania* and *fuqarotania* ‘whisper about’. The conjugation analysis could easily accommodate cases of this sort by saying that certain bases permit more than one variant of a given suffix, in free variation.

However, there are also cases where there is a semantic difference between the multiple forms derived from the same base, as in Toqabaqita *ngatalia* ‘speak about, discuss’ and *ngatafia* ‘speak angrily to, berate’ (*ngata* ‘speak’); and Manam *toaqi* ‘imitate, copy’ and *tongaqi* ‘teach s.o. (to do s.th.)’ (*to* ‘try [to do s.th.]’). If Toqabaqita *-li* and *-fi* (and the other forms) were co-allomorphs and, similarly, if Manam *-aq* and *-ngaq* (and the other forms) were co-allomorphs, one would not expect them to give rise to different meanings.

#### 5.5 Object-indexing and other suffixes in Manam

The next argument against the conjugation analysis is specific to Manam. Manam has several types of suffixes added to transitive verbs. One of these is a set of suffixes whose function is to index direct objects. There are seven basic object suffixes, for first, second and third person, singular and plural, and an exclusive–inclusive distinction in the first person plural. (Some of the suffixes have more than one allomorph, but their distribution is not relevant to the issue at hand.) The same set of object suffixes is used with both short and long transitives. What is of relevance here is the fact that with a given verb the same thematic consonant is used with all the object suffixes. Or, put differently, for a given transitive verb all the object suffixes are associated with the same thematic consonant. It is never the case that for a given transitive verb a certain thematic consonant is used with some but not all the object suffixes, while another thematic consonant is used with (some of) the other object suffixes. (But, of course, multiple transitive verbs derived from the same base may use different thematic consonants.) Table 3 gives a complete paradigm of the object suffixes with a short transitive verb, and a partial paradigm with a long transitive verb. (Verbs also take subject-indexing prefixes; these have been omitted for simplicity’s sake. The initial *i* in the plural, but not the singular, object suffixes is epenthetic.) If the thematic consonants were part of the object suffixes, there would be no obvious reason why they should be the same in all the object suffixes for a given verb. (This would not be an issue on the phonological analysis.)

Table 3: Object-indexing suffixes in Manam

<i>longo-r-a</i>	'hear me'
hear-TC-1SG:O	
<i>longo-r-iqo</i>	'hear you (singular)'
<i>longo-r-i</i>	'hear him/her/it'
<i>longo-r-iqama</i>	'hear us (exclusive)'
<i>longo-r-iqita</i>	'hear us (inclusive)'
<i>longo-r-iqaming</i>	'hear you (plural)'
<i>longo-r-idi</i>	'hear them'
<i>rere-t-aq-a</i>	'like me'
like/want-TC-TR-1SG:O	
<i>rere-t-aq-iqo</i>	'like you (singular)'
<i>rere-t-aq-iqaming</i>	'like you (plural)'
<i>rere-t-aq-idi</i>	'like them'

In long transitives, the thematic consonant and (the rest of) the object suffix are separated by the transitive suffix. One would have to argue either that *aq* is an infix, or that with long transitive verbs the thematic consonant is part of the transitive suffix, not of the object suffix. In the latter case, each object suffix would have a number of allomorphs, one without a thematic consonant, used with long transitives (the thematic consonant being part of the transitive suffix), and a number of allomorphs with thematic consonants used with short transitives. The transitive suffix also would have a number of allomorphs: *-taq*, *-qaq*, *-maq* and so on (cf. Table 2). Clearly, this kind of analysis would result in great complexity.

Manam has a nominalising morpheme suffixed to verbs, transitive or intransitive. Examples (47) and (48) illustrate nominalisations of transitive verbs with thematic consonants, a short transitive and a long transitive, respectively. (The intransitive verb 'urinate' is *tamim*; the second *i* in (48) is epenthetic.)

- (47) *gere-t-a*  
tattoo/draw/write-TC-NOM  
'tattooing, drawing, writing'
- (48) *tamimi-r-aq-a*  
urinate-TC-TR-NOM  
'urinating of s.th. out, passing of s.th. in urine'

Two possibilities arise. Either the thematic consonant is part of the nominalising suffix, which means that *aq* is infixed into the nominaliser; or both the transitive suffix and the nominalising suffix may contain thematic consonants, the latter only with short transitives. In either case, the result would be great complexity.

Furthermore, for a given transitive verb the thematic consonant used in the nominalisation is the same as the one used with the object suffixes. Compare *gere-t-a* 'tattooing, drawing, writing' in (47) and *gere-t-iqo* 'tattoo you (SG)', and *tamimi-r-aq-a* 'urinating of s.th. out, passing of s.th. in urine' in (48) and *tamimi-r-aq-i* 'urinate s.th. out, pass s.th. in urine'. On the conjugation analysis, this would be a mere accident. (This too would not be an issue on the phonological analysis.)



Manam also has a suffix *-i* used to express repetition of a sequence of two or more events on a given occasion, typically a relatively rapid sequence of events (Lichtenberk 1983:253). The sequential events are encoded by nominalised verbs, as in (49).

- (49) *Niu qabu-n-i-a qozo-m-i-a*  
 coconut heap-TC-REP-NOM husk-TC-REP-NOM
- ota-q-i-a elua-q-i-a...*  
 break-TC-REP-NOM carry-TC-REP-NOM
- 'As soon as [we] collected [lit. 'heaped up'] a few coconuts, we husked them, broke them, and carried them away, collected some more, husked them, broke them, and carried them away, ...'

The thematic consonants would be part either of the repetition morpheme or of the nominaliser, in which case different kinds of complexity would arise, parallel to those discussed above. Furthermore, for a given transitive verb the thematic consonant used in the repetition–nominalisation form is the same as that used in the plain nominalisation and with the object suffixes; for example, the repetition–nominalisation *qozo-m-i-a* 'repetitive husking', *qozo-m-a* 'husking' and *qozo-m-i* 'husk it'. This is powerful evidence against the conjugation analysis.

## 6 An alternative analysis

In §4 and §5 respectively, a number of arguments were presented against the phonological and the conjugation analyses of the thematic consonants in Toqabaqita and in Manam. In some, but only in some, cases, evidence against one type of analysis could be interpreted as evidence in favour of the other type. For example, the fact that in Manam for a given transitive verb the thematic consonant is the same with all the object suffixes and with the two types of nominalisation presents great problems for the conjugation analysis, but on the phonological analysis the problem would simply not arise. On the other hand, the fact that among lexical morphemes only verbs could end in a consonant in Toqabaqita and in a consonant other than *m* and *ng* in Manam speaks against the phonological analysis and in favour of the conjugation analysis. Variation in thematic consonants is problematic for both types of analysis, in different ways.

In what follows, I will propose an alternative analysis of the thematic consonants in the two languages, one that avoids the drawbacks both of the phonological and of the conjugation analyses. I will refer to it as the **morpholexic analysis**. Briefly, on the morpholexic analysis the thematic consonants are separate morphs, but they are semantically empty morphs. Of themselves, they do not carry any meaning; they are not signs. Rather, it is only the combination of the (inner) base, the thematic consonant, and, in some cases, a transitive suffix that is meaningful. It is only the lexemes, the transitive verbs as wholes, that are signs. I will now discuss the morpholexic analysis in more detail.

As the discussion of the conjugation analysis in §5.2 showed, the thematic consonants are isolatable as separate morphs because of their recurrence. Since they are not part of the inner base, none of the arguments against the phonological analysis applies; and since they are not part of the affixes, none of the arguments against the conjugation analysis applies either. In the lexicon, there will be 'redundancy' rules linking transitive verbs and the relevant bases. More than one transitive verb may be linked to one and the same base. The thematic

consonants being semantically empty, there is no conflict if more than one thematic consonant is associated with one and the same base, even when the derived forms are different in meaning. The thematic consonants in short and long transitives are drawn from the same pool, which automatically accounts for the relations between the sets used in short transitives and the sets used in long transitives (as discussed in §5.3), the ranges of variation simply being smaller in long transitives.

The account of the thematic consonants in Toqabaqita and in Manam presented here falls squarely within lexeme-based morphology, as developed by Aronoff (1976, 1994)—see also Anderson (1992), who speaks of lexical-stem formation rather than lexeme formation; and also Matthews (1991). In Aronoff's view of morphology, "the morpheme is not the basic unit of language" (1994:9); rather, the lexeme is. Let us return to some of the data we saw earlier in order to justify this statement.

Toqabaqita has an intransitive verb *ngata* 'speak', on which two transitive verbs are based: *ngata-l-i-a* 'speak about, discuss' and *ngata-f-i-a* 'speak angrily to, berate'. One cannot attribute the semantic difference between the two transitive verbs to different meanings carried by the thematic consonants. Both *l* and *f* are found as thematic consonants in many other transitive verbs, but they cannot be said to contribute more-or-less stable meanings. One finds *l* in, for example, *busu-l-i-a* 'cause to burst' (*busu* (intr) 'burst'); and *fai-l-i-a* 'clean, clear (a garden) by sweeping stones, sticks, leaves with one's hands', (*fai* (intr) 'scrape, scratch'). And one finds *f* in, for example, *kulu-f-i-a* 'cause to bend down' (*kulu* 'hang down') and *lae-f-i-a* 'go for, go in order to get s.th.' (*lae* 'go'). (See §7 for more discussion of thematic *f*.) Note also from the examples just given that the relation between a transitive verb and its base may be causative or 'applicative'. The thematic consonants do not have any stable, identifiable semantic content. (However, as will be discussed in §7, there are some partial associations between thematic consonants and the meanings of the verbs in which they occur.)

With respect to transitive verbs such as those mentioned in the preceding paragraph, there are two possibilities concerning what constitutes the minimal sign: (i) the combination of the (inner) base and the thematic consonant, as in *ngata-l*; or (ii) the combination of the (inner) base, the thematic consonant and the transitive suffix, as in *ngata-l-i*. The latter possibility is more viable because there may be both a short transitive verb and a long transitive verb derived from the same base that have the same thematic consonant but are distinct in meaning: witness *liu-f-i-a* 'walk, roam around (a place)' and *liu-f-ani-a* 'befall, "visit" (of a disease or death befalling a person as retribution for taboo behaviour)' (*liu* 'walk, roam'). Thematic consonants as empty morphs in Toqabaqita and in Manam are reminiscent of the theme vowels in Latin, also empty morphs (Aronoff 1994).

There is one more matter that needs to be addressed. Toqabaqita has a short transitive suffix *-i* and, on the present analysis, long transitive suffixes *-aqi* and *-ani*. Since the thematic consonants are treated as separate morphs, why not treat the short suffix *i* and the final *i*'s in the long suffixes as morphs? The answer is that nothing would be gained by doing that. Analysing the thematic consonants as separate morphs is motivated; not doing so would result in a number of analytical problems, as the discussion of the phonological and the conjugation analyses demonstrated. On the other hand, nothing would be gained by considering the final *i* in *-aqi* and *-ani* to be a separate morph or morpheme. Such an analysis would be completely unmotivated, positing internal structure where there is no independent evidence of it.

## 7 Morphs that are empty but not without any semantic associations

The thematic consonants in Manam and in Toqabaqita are empty morphs: they do not provide independent contribution to the meanings of transitive verbs. One might then expect the choice of a thematic consonant to be completely arbitrary: given the meaning of a base, any of the thematic consonants would be as likely to occur in the derived form as any other. This, however, is not entirely true.

It has been noted by several researchers that in at least some Oceanic languages the choice of a thematic consonant is not entirely haphazard, that certain meanings of bases tend to be associated with certain thematic consonants; and, conversely, that at least some thematic consonants tend to go with bases of certain meanings. Arms (1974) has postulated for (Standard) Fijian various correlations between the individual thematic consonants and the meanings of the bases with which they occur or tend to occur. Similarly, I have suggested that the distribution of the thematic consonants in Manam is not entirely arbitrary (Lichtenberk 1978). However, the approaches adopted by Arms and by myself are rather different. Arms tended to look for an all-embracing meaning for each thematic consonant, which required positing very general semantic categories. For example, he says that thematic *c* is associated with the meaning 'pliancy, gentle contact, bodily experience', and he characterises the semantic force of thematic *n* as 'X-ify, en-, be-, form, support'. Even so, there are many exceptions. In contrast, I did not necessarily look for unitary, general meanings to be associated with the individual thematic consonants in Manam; rather, the pattern I saw was one where there were more specific semantic fields associated with the individual thematic consonants, with more than one such field per consonant. The approach to thematic consonants in terms of multiple semantic clusterings rather than some vague general meanings has also been adopted by Geraghty (1983) for Fijian and by Clark (1977) for Fijian, Sa'a and Nggela. For example, while Arms associates thematic *c* with 'pliancy, gentle contact, bodily experience', I see Manam thematic *r*, which historically corresponds to Fijian *c*, as associated with at least three semantic fields: (i) excretion, secretion; (ii) wetting, moistening, immersing; and (iii) stirring, mixing, gathering. However, there are exceptions in both directions in Manam. First, a verb that belongs in one of the three categories does not have *r* as the thematic consonant (see *suburawa-i* 'sweat on' and *suburawa-ng-aq-i* 'sweat out' in Table 4). And second, thematic *r* is also used with verbs that do not belong in any of the three categories, such as *ado-r-i* 'make straight' (*ado* 'be straight'). In the Manam sample, out of seventy-seven instances of *q* as a thematic consonant, forty-one were verbs referring to 'destructive' actions, such as breaking, hitting, chopping and cutting (short and long transitive verbs both being counted). But there were also thirty instances of no thematic consonant with verbs of such meanings, twelve instances of *t*, seven of *ng*, three of *n* and three of *l*. Thematic *q* was not infrequently also used with verbs of translational and nontranslational motion.

One finds some semantic correlations between thematic consonants and verb meanings in Toqabaqita as well. Verbs of excretion and secretion are strongly associated with thematic *s* in short transitives and thematic *t* in long transitives (the *s:t* correspondence being the result of an earlier sound change). Table 4 contains a list of verbs of secretion and excretion in Manam and in Toqabaqita: in Manam *r* is dominant, and in Toqabaqita *s* and *t* are dominant. Now, Manam *r* on the one hand and Toqabaqita *s* and *t* on the other are in historical correspondence with each other, which suggests that this particular association between meanings and a thematic consonant goes back in time. Similarly, Fijian *c* historically corresponds to Manam *r* and Toqabaqita *s* and *t*; and in Fijian, verbs of excretion and

secretion (included in Arms' category of bodily experience) are strongly associated with thematic *c*. On present subgrouping hypotheses, Manam, Toqabaqita and Fijian belong in different primary subgroups of Oceanic, which suggests that this particular link between meanings and a thematic consonant began to develop already in Proto Oceanic times.

**Table 4:** Verbs of secretion and excretion in Manam and in Toqabaqita

	SHORT TRANSITIVES	LONG TRANSITIVES
Man	<i>tamimi-r-i</i> 'urinate on'	<i>tamimi-r-aq-i</i> 'urinate out'
Toa	<i>kwara-s-i-a</i> 'urinate on'	<i>kwara-t-ani-a</i> 'urinate out'
Man	<i>tabeqa-r-i</i> 'defecate on'	<i>tabeqa-r-aq-i</i> 'excrete out'
Toa	<i>feqe-s-i-a</i> 'defecate on'	<i>feqe-t-ani-a</i> 'excrete out'
Man	<i>qulena-r-i</i> 'vomit on'	<i>qulena-r-aq-i</i> 'vomit out'
Toa	<i>moa-s-i-a</i> 'vomit on'	<i>moa-t-ani-a</i> 'vomit out'
Man	<i>mongo-r-i</i> 'spit at'	<i>mongo-r-aq-i</i> 'spit out'
Toa	<i>ngisu-f-i-a</i> 'spit at'	<i>ngisu-t-ani-a</i> 'spit out'
Man	<i>qongi-r-i</i> 'soil with nasal mucus'	<i>qongi-r-aq-i</i> 'cause (mucus) to come out of one's nose'
Toa	no evidence of transitive forms	
Man	<i>puqi-r-i</i> 'pass wind at'	<i>puqi-r-aq-i</i> 'emit (smell) when passing wind'
Toa	no evidence of transitive forms	
Man	<i>suburawa-i</i> 'sweat on'	<i>suburawa-ng-aq-i</i> 'sweat out'
Toa	no evidence of transitive forms	

One would not be surprised to find that cognate verbs contain the corresponding thematic consonants. Thus, it is to be expected that the transitive reflexes of Proto Oceanic \*bekas 'excrete' (Ross 1988) in Manam, Toqabaqita and Fijian contain *r*, *s/t* and *c* respectively: they all reflect Proto Oceanic \*s (see Table 4 and note also Fijian *veka-c-a* 'excrete on' and *veka-c-ak-a* 'excrete out'). However, the same is also true of synonymous verbs that are not cognate, such as Manam *qulena-r-i*, Toqabaqita *moa-s-i-a* and Fijian *lua-c-a*, all of which mean 'vomit on'. It appears that when the original verb 'vomit' was replaced in the history of a language, the semantically 'appropriate' thematic consonant was used in the replacement.

There is another identifiable cross-linguistic pattern. In Toqabaqita, verbs of motion exhibit association with *f* as the thematic consonant in short transitives. The object of the transitive verb is a goal, destination, or the place in, into, through which the motion takes place, for example *lae-f-i-a* 'go for, go in order to get s.th.', *ango-f-i-a* 'crouch "for"' (e.g. crouch "for" a bird, as a hunter may do when trying to shoot a bird), *olo-f-i-a* 'of a means of transportation: arrive at (a place)', *kotho-f-i-a* 'go into, enter', *qidu-f-i-a* 'shuffle on one's bottom on' (e.g. a baby shuffling on the ground), but *lofo-q-i-a* 'pounce on', not \**lofo-f-i-a* (*lofo* 'fly, jump'). Similarly, Clark (1977) has found an association between verbs of motion and thematic *h* in Sa'a and verbs of motion and thematic *v* in Nggela. Sa'a and Nggela are relatively closely related to Toqabaqita, and Toqabaqita *f*, Sa'a *h* and Nggela *v* reflect the same earlier sound. However, as pointed out by Clark (1977) and by Geraghty (1983), a similar association is also found in Fijian with *v* as the thematic consonant, which corresponds to Toqabaqita *f*, as in *cici-v-a* 'run to or for', *qalo-v-a* 'swim to (a place)', *lako-v-a* 'go to or through' (Capell 1973). Since very much the same pattern is found with short transitives in

Toqabaqita (and Sa'a and Nggela) and in Fijian, some such association between verb meanings and a thematic consonant most likely existed already in Proto Oceanic. (No such pattern is evident in Manam, but it needs to be noted that Toqabaqita *f* and Fijian *v* usually correspond to Manam zero.) A suggestion that there was a link already in Proto Oceanic between certain thematic consonants and certain meanings—specifically verbs of secretion/excretion and verbs of motion—has been made by Geraghty (1983). If some such links did indeed exist, the thematic consonants may have gravitated, or may have begun to gravitate, away from the bases already in Proto Oceanic times, assuming that associations between thematic consonants and semantic fields via analogical extensions are more likely to develop if the consonants are not part of the (inner) base.

There are also instances of associations between thematic consonants and semantic fields that appear to be language specific. One that is particularly striking is found in Manam. Manam has a number of short transitive verbs based on nominal kinship terms and terms for friends and partners. The meaning of a derived verb based on a noun *X* is something like 'have/regard/treat somebody as one's *X*'. All of the forms of which I am aware have *m* as their thematic consonant (see Table 5). There is no evidence for reconstructing this pattern for Proto Oceanic, even though most of the Manam nouns do continue Proto Oceanic forms.

**Table 5:** Manam transitive verbs based on kinship and like nouns

<i>tama-m-i</i>	'regard s.o. as one's father (behave to that person as if he were one's genetic father, e.g. when adopted by him)'
<i>tina-m-i</i>	'regard s.o. as one's mother'
<i>toqa-m-i</i>	'regard s.o. as one's older sibling of same sex'
<i>tari-m-i</i>	'treat, adopt s.o. as one's younger sibling of same sex'
<i>natu-m-i</i>	'adopt s.o. as one's child'
<i>taua-m-i</i>	'have s.o. as one's trading partner'
<i>ruanga-m-i</i>	'have s.o. as one's friend, befriend s.o.'

However, even though the verbs based on kinship terms all take *m* as their thematic consonant, it is not the case that *m* is used only with these verbs, (e.g. *nodo-m-i* 'worry about, feel sad, sorry on account of', *raru-m-i* 'scoop up', *zoqu-m-i* 'bend, fold'). In fact, the verbs based on kinship terms form only a minority of the set of verbs that take *m* as their thematic consonant.

## 8 Concluding remarks

There is evidence both in Toqabaqita and in Manam against analysing the thematic consonants as part of the inner base, and there is also evidence against analysing them as part of the transitive or object suffixes. The thematic consonants are rather morphs of their own, but they are empty morphs. At the same time, at least some of the thematic consonants are not completely 'neutral' semantically. With verbs that belong in a given lexical field, a certain thematic consonant may be more likely to occur than any of the other consonants. And in both languages there are cases where pairs of transitive verbs that are different in meaning, but derived from the same base, are only distinguished by their thematic consonants. True, but from these facts it does not follow that the thematic consonants have meanings of their own. The associations between semantic fields and thematic consonants are too weak;

there are far too many cases where a verb that belongs in a given lexical field does not take the 'expected' thematic consonant, and the semantic differences between transitive verbs and their bases are not stable. In cases where a difference in thematic consonants is associated with a difference in meaning in verbs derived from the same base, it does not make sense to say that the difference in meaning is **due to** the different thematic consonants. Thus, it does not make sense to say that the difference between Toqabaqita *ngata-l-i-a* 'talk about, discuss' and *ngata-f-i-a* 'speak angrily to, berate' is due to different semantic contributions of *l* and *f*. Cases like these are not very common, and there is no consistency.

The thematic consonants are analysable as morphs only on formal grounds, not on semantic grounds; they are not signs. It is only the combination of a base, a thematic consonant and/or a transitive suffix that is meaningful. Lexemes, as signs, may be combinations of signs and nonsigns, something explicitly recognised by a lexeme-based view of morphology (Aronoff 1994). To the extent that there are some associations between thematic consonants and semantic fields, it looks as though the thematic consonants were morphs in search of meanings.

## References

- Anderson, Stephen R., 1992, *A-morphous morphology*. Cambridge: Cambridge University Press.
- Arms, David G., 1974, *Transitivity in Standard Fijian*. PhD dissertation, University of Michigan.
- Aronoff, Mark, 1976, *Word formation in generative grammar*. Linguistic Inquiry Monograph 1. Cambridge, Mass.: MIT Press.
- 1994, *Morphology by itself: stems and inflectional classes*. Cambridge, Mass.: MIT Press.
- Blevins, Juliette, 1994, A phonological and morphological reanalysis of the Maori passive. *Te Reo* 37:29–53.
- Capell, A., 1973, *A new Fijian dictionary* (4th edn). Suva: Government Press.
- Clark, Ross, 1977, Eastern Oceanic transitive suffixes and the genesis of rules. Paper presented at the Austronesian Symposium, Linguistic Society of America Institute, University of Hawai'i.
- Geraghty, Paul A., 1983, *The history of the Fijian languages*. Oceanic Linguistics Special Publication No. 19. Honolulu: University of Hawai'i Press.
- Hale, Kenneth, 1973, Deep-surface canonical disparities in relation to analysis and change: an Australian example. In Thomas A. Sebeok, ed. *Current trends in linguistics*, 11:401–458. The Hague: Mouton.
- 1991, Remarks on G. Sanders 'Levelling in the history of Polynesian passive formations'. *Journal of the Polynesian Society* 100:99–101.
- Hohepa, Patrick W., 1967, *A profile generative grammar of Maori*. Memoir 20. Bloomington: Indiana University Publications in Anthropology and Linguistics.

- Kaye, Jonathan, 1975, A functional explanation for rule ordering in phonology. In Robin E. Grossman, L. James San and Timothy J. Vance, eds *Papers from the Parasession on Functionalism*, 244–252. Chicago: Chicago Linguistic Society.
- Lichtenberk, Frantisek, 1978, Thematic consonants in Manam transitive verbs. *Anthropological Linguistics* 20:185–193.
- 1983, *A grammar of Manam*. Oceanic Linguistics Special Publication No. 18. Honolulu: University of Hawai'i Press.
- 1988, The Cristobal-Malaitan subgroup of Southeast Solomonian. *Oceanic Linguistics* 27:24–62.
- Matthews, P.H., 1991, *Morphology* (2nd edn). Cambridge: Cambridge University Press.
- McCarthy, John J., 1981, The role of the evaluation metric in the acquisition of phonology. In C.L. Baker and John J. McCarthy, eds *The logical problem of language acquisition*, 218–248. Cambridge, Mass.: MIT Press.
- Pawley, Andrew, 1972, On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol. 3, 1–142. Pacific Anthropological Records No. 13. Honolulu: Bernice Pauahi Bishop Museum.
- 1973, Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103–188.
- Ross, M.D., 1988, *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.
- Sanders, Gerald, 1990, On the analysis and implications of Maori verb alternations. *Lingua* 80:149–196.
- 1991, Levelling and reanalysis in the history of Polynesian passive formations. *Journal of the Polynesian Society* 100:71–90.





# *The development of morphologically complex possessive markers in the Southern Vanuatu languages*

---

JOHN LYNCH

Although not phonologically ‘well-behaved’, Southern Vanuatu languages nevertheless contain a large number of lexical items clearly derivable from Proto Oceanic reconstructed forms, and their grammars are fairly Oceanic-looking. However, there have been a number of innovations, among them a great diversity in the shape of the indirect possessive markers, despite possessive forms and structures that otherwise clearly reflect Proto Oceanic antecedents. This study examines innovations in the indirect possessive markers in Southern Vanuatu languages, with particular attention to the processes by which the originally monomorphemic Proto Oceanic markers have acquired extra morphological material, and yet have ended up monomorphemic once again.

## **1 Introduction<sup>1</sup>**

The languages of Southern Vanuatu constitute a closed subgroup of Oceanic. Although not phonologically ‘well-behaved’, a large number of lexical items in each of these languages are nevertheless clearly derivable from Proto Oceanic (POc) reconstructed forms, and their grammars are fairly Oceanic-looking, although there have been a number of innovations. A particularly puzzling issue, however, is the diversity in the shape of the indirect possessive markers, despite the fact that the overall possessive system and the shape of the pronominal suffixes continue Proto Oceanic forms and structures.

This study examines some of the changes that the Southern Vanuatu (SV) languages have made to the original Proto Oceanic possessive-marking system. Some of these, like the loss

---

<sup>1</sup> I am pleased to be able to dedicate this paper to Byron Bender. As a scholar, his influence on Micronesian linguistics has been considerable. As a teacher and longtime department chair, his influence on graduate students at the University of Hawai‘i over the past thirty years has been immense. I am also grateful to Joel Bradshaw, Terry Crowley, Robert Early and Malcolm Ross for comments on an earlier draft of this paper.

of the widespread POC GENERAL possessive marker \*na-, are relatively trivial in any broad morphological context, though of some interest to Oceanists. One feature of rather broader interest, however, is that most monomorphemic Proto Oceanic markers were converted into bi- or even trimorphemic forms, though by a variety of different processes, and were then subsequently reanalysed as simple monomorphemic forms. I will discuss this in some detail at the end of this paper.

## 2 Proto Oceanic and Southern Vanuatu possessive systems

Proto Oceanic is reconstructed (Lynch 1996) as having had the possessive-marking system as set out in (1), where I also list the possessive markers which can be reconstructed for Proto–Southern Vanuatu (PSV), on the basis of the data in Table 1.<sup>2</sup>

(1)	PROTO OCEANIC	PROTO SOUTHERN VANUATU
KIN	Direct possession	Direct possession
PART	Direct possession	Direct possession
PASSIVE	Direct possession	*ra, ira-
FOOD	*ka-	*na-ya-
DRINK	*ma-	*[n]-m <sup>w</sup> a-
PLACE		*(i,u)-i(u)m <sup>w</sup> a-
GENERAL	*na- or *a- (and possibly also *ta- and/or *sa-)	See §4 below

There is some evidence (Lynch, Ross & Crowley, forthcoming) that, in addition to the possessive markers listed in (1), Proto Oceanic also had, or was beginning to develop, a system of possessive classifiers that derived fairly directly from nouns.

With pronominal possessors, a pronominal suffix was attached directly to the noun in cases of direct possession, and to the possessive marker/classifier in indirect possession; while with (nonspecific?) nominal possessors, the directly possessed noun or the possessive marker/classifier took a construct suffix. The Proto Oceanic (Lynch, Ross & Crowley, forthcoming) and Proto Southern Vanuatu suffixes were as listed in (2); the nonsingular Proto Southern Vanuatu forms were further suffixed by markers of number.

(2)	PROTO OCEANIC	PROTO SOUTHERN VANUATU
1SG	*-gu	*-g
2SG	*-mu	*-m[u]
3SG	*-ña	*-n[i]
1INC.PL	*-da	*-d-
1EXC.PL	*-ma[m]i	*-m[am]-
2PL	*-m[i]u	*-mi-
3PL	*-dra	*-nr-
CONSTRUCT	*-qi or *-i	*-i

<sup>2</sup> Small capitals are used to designate categories of possession; the semantics of the more common of these will, in general, not be discussed here. Data in languages are cited in phonemic rather than standard orthographic form (except that Anejom̄ *j* represents /tʃ/). Abbreviations of grammatical categories: AOR – aorist, ART – article, CONT – continuous, CS – construct suffix, EX – exclusive, GEN – general, INC – inclusive, LOC – locative, OBL – oblique, PL – plural, POSS – possessive, RECPST – recent past, SG – singular.

**Table 1:** Possessive markers and suffixes in Southern Vanuatu languages

MARKERS	ERROMANGO			TANNA					ANEITYUM
	SYE	URA	UTAHA	NORTH TANNA	WHITESANDS	LENAKEL	S.W. TANNA	KWAMERA	ANEJOM
KIN	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct
PART	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct	Direct
PASSIVE	<i>ra, ira-</i>	<i>ra, ira-</i>	?	<i>e, (?)</i>	<i>ia, la-</i>	<i>le, la-</i>	<i>ie, ila-</i>	<i>ia, ira-;</i> <i>{kəfu- ~ kə}</i>	<i>a, era-</i>
FOOD				<i>nəŋa-</i>	<i>nəŋə</i>	<i>nəkə</i>	<i>na-</i>	<i>sana-</i>	<i>nəŋa-</i>
DRINK				<i>nəm<sup>w</sup>ə</i>	<i>nəm<sup>w</sup>ə</i>	<i>nən<sup>w</sup>ə</i>	<i>{ni-}</i>	<i>sanm<sup>w</sup>u-</i>	<i>lum<sup>w</sup>a-</i>
PLACE				<i>iim<sup>w</sup>a-</i>	?	<i>iim<sup>w</sup>a-</i>	<i>iim<sup>w</sup>a-</i>	<i>im<sup>w</sup>a-</i>	<i>um<sup>w</sup>a-</i>
PLANT				<i>nai-</i>	<i>nai-</i>	<i>ne-</i>	<i>ni-</i>	<i>{sap<sup>w</sup>as-}</i>	
JUICE									<i>{liŋa-}</i>
GENERAL	<i>hore-</i> <i>horo-</i> <i>ihe-</i>	<i>ar(V)-</i>	<i>et(V)-</i>	<i>raha-</i>	<i>raha-</i>	<i>taha-</i>	<i>kapa- ~ kafa-</i> <i>kapaha-</i>	<i>sava-</i>	<i>u- ~ uwu-</i>
SUFFIXES									
1SG	<i>-ŋ</i>	<i>-k</i>	<i>-ŋ(or Ø?)</i>	<i>-k</i>	<i>-k</i>	<i>-k</i>	<i>-k</i>	<i>-k</i>	<i>-k</i>
2SG	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m</i>	<i>-m<sup>w</sup></i>
3SG	<i>-n</i>	<i>-n</i>	<i>-n</i>	<i>-n</i>	<i>-n</i>	<i>-n</i>	<i>-n, -ni</i>	<i>-n, -ni</i>	<i>-n</i>
1INC:NONSG	<i>-t ~ -nt</i>	?	?	<i>-t</i>	<i>-t</i>	<i>-t</i>	<i>-t</i>	<i>-t</i>	<i>-j</i>
1EXC:NONSG	<i>-mam</i>	?	<i>{-kum}</i>	<i>{-tm-}</i>	<i>{-təm-}</i>	<i>-m-</i>	<i>-m-</i>	<i>-m-</i>	<i>-m-</i>
2NONSG	<i>-mi</i>	?	<i>-mim</i>	<i>{-təm-}</i>	<i>{-təm<sup>w</sup>}</i>	<i>-mi-</i>	<i>-mi-</i>	<i>-mi-</i>	<i>-mi-</i>
3NONSG	<i>-nr</i>	?	<i>-ra</i>	<i>-l-</i>	<i>-l-</i>	<i>-nil-</i>	<i>-li-</i>	<i>-nr-</i>	<i>-r-</i>
CONSTRUCT	<i>{-n}</i>	<i>{-n}</i>	?	<i>-ə ~ Ø</i>	<i>-ə Ø</i>	<i>-ə Ø</i>	<i>-i</i>	<i>-i</i>	<i>-i</i>

NOTES:

1. The forms given for the Erromangan languages represent the probable precontact situation (see Crowley, this volume, for more recent changes): Utaha is now extinct and Ura is moribund. The Ura possessive suffixes listed here are vestigial in the modern language.
2. Blanks indicate that the category does not exist; ? indicates that the category probably exists (or existed) but has not been recorded.
3. Other than under GENERAL, I have used braces {...} to mark forms that do not reflect a PSV or Proto Tanna reconstruction.
4. Some morphophonemic alternants are shown, but in many cases I have ignored these here.
5. The two GENERAL markers in Southwest Tanna represent different dialects.
6. Hyphens after the nonsingular Tanna and Anejom pronominal suffixes indicate that number suffixes must be added.

The Erromangan languages have retained only KIN/PART, PASSIVE and GENERAL marking from the original system, and have lost all other markers (see Crowley, this volume, on these and other more recent developments in the Erromangan systems). The Tanna languages and Anejoñ, on the other hand, have not only retained all the Proto Oceanic categories, but have each added two additional ones: they both show a place category (which can be reconstructed for Proto Southern Vanuatu); Proto Tanna has also added a plant category (referring to trees or plants which the possessor has planted), and Anejoñ a JUICE category (referring to possessions from which the juice is sucked).

### 3 Passive possession

Before coming to the main focus of this paper, it is perhaps worth spending a little time on one other interesting development in the Southern Vanuatu languages. The term ‘passive possession’ is used to describe a relationship between the possessor and something done to or about him or her: thus ‘my wound’, ‘his story (the one told about him)’, ‘the pig’s (tether) rope’, and ‘her slap (the one she received)’ are all prototypical examples of passive possession. This category was probably marked by a direct construction in Proto Oceanic; but there is a very widespread tendency within Oceanic for this category to be marked by an indirect construction with \*ka-,<sup>3</sup> and a less widespread tendency for it to be subsumed under the GENERAL category.

In Southern Vanuatu languages, however, none of these strategies operates. Rather, the expression of passive possession came to be marked by the general oblique preposition PSV \*ra which marks, among other things, time, location, direction, instrument, cause and reference. It had two grammatically conditioned allomorphs—\*ra governing nouns and \*ira- governing pronouns—which then occurred in their possessive suffix forms (the exact conditioning of the reflexes of \*ra and \*ira- varies slightly from one modern Southern Vanuatu language to another, and need not concern us here); for example:

(3) SYE

*Yoʔoʔ-velom iŋko ra naʔave.*  
 1SG:RECPST-come here OBL kava  
 ‘I have come here for kava.’

*Yoʔoʔ-velom iŋko ira-m.*  
 1SG:RECPST-come here OBL-2SG:POSS  
 ‘I have come here for you.’

(4) LENAHEL

*R-am-arək le nəkɪnhamra.*  
 3SG-CONT-live OBL bush  
 ‘He lives in the bush.’

<sup>3</sup> The explanation for this widespread tendency is less clear. It may have to do with the formal similarity of the possessive marker \*ka- and the Proto Western Oceanic benefactive—or, better, affective—preposition \*ka. It may also, or instead, have to do with the fact that POc \*kani ‘eat’ may have had a secondary meaning ‘suffer, experience (negatively)’, since in a number of Oceanic languages this verb is used not only for eating food but also for suffering ‘the slings and arrows of outrageous fortune’; and thus both kinds of nouns which are objects of \*kani are possessed with \*ka- in many Oceanic languages.

*R-am-eiua la-k.*  
 3SG-CONT-lie OBL-1SG:POSS  
 'He's lying to me.'

- (5) ANEJOŃ  
*Et yip<sup>w</sup>al a np<sup>w</sup>oθeθ a nuyaleŋ aan.*  
 3SG:AOR tell.story OBL bush.spirit OBL morning he  
 'He told a story about a bush spirit in the morning.'

*Et yip<sup>w</sup>al era-k aan.*  
 3SG:AOR tell.story OBL-1SG:POSS he  
 'He told a story about me.'

This same morpheme marks passive possession, which category includes not only things done to or about the possessor, but also certain part terms and, in some languages, certain kin terms:<sup>4</sup>

- (6) SYE  
*tali ira-m* 'your shadow'  
*uvuvu ira-n natmonuγ* 'a story about a/the chief'  
*neipolu ra (~ ira-n) nompyahi* 'the pig's (removed) stomach'
- (7) LENAHEL  
*nouanaje la-k* 'my story (told about me)'  
*uulpəs le kuri* 'the dog's heel'
- (8) ANEJOŃ  
*nyip<sup>w</sup>al era-m<sup>w</sup>* 'your story (told about you)'  
*nataheñ era-i Tanipe* 'Tangipe's sister'  
*nalau a ntijja-n* 'the hair of his ears'

#### 4 Nominalisation of possessive markers

The other possessive markers in the Southern Vanuatu languages are all synchronically monomorphemic, but nearly all are, historically at least, bimorphemic, with many of the Kwamera ones being historically trimorphemic. This section will deal with the markers of all except the GENERAL category; I will ignore the *sa-* prefix on these markers in Kwamera until §6.

##### 4.1 Locativisation

The Tanna languages and AnejoŃ have a locational possessive marker marking the possession as a place, generally in the sense of one's home village or an area of land or sea which is owned according to traditional custom by the possessor. Contrast the following:

<sup>4</sup> This category is referred to as "semi-alienable possession" in various earlier works on Tanna languages and as "removed inalienable possession" in Crowley's grammar of Sye (1998:171–172). See also Lynch (1992) for a discussion of the use of this construction with certain part terms.

- (9) LENAHEL  
*nauanu iim<sup>w</sup>a-k*  
 village POSS.PLACE-1SG  
 ‘my village (where I was born, where I have rights)’  
*nauanu taha-k*  
 village POSS.GEN-1SG  
 ‘my village (the village where I am currently living)’
- (10) ANEJOŨ  
*naworitai um<sup>w</sup>a-i natimariθ*  
 garden POSS.PLACE-CS chief  
 ‘the chief’s (traditionally owned) garden-land’  
*naworitai u natimariθ*  
 garden POSS.GEN chief  
 ‘the chief’s garden (not on his own land or, if it is, this is not pragmatically important)’

The marker clearly derives from POC \*Rum<sup>w</sup>aq, PSV \*n-i(u)m<sup>w</sup>a ‘house’—compare with Lenakel *nim<sup>w</sup>a*, AnejoŨ *niom<sup>w</sup>* ‘house’, which contain the fused article (here just *n*). The form in most of the Tanna languages is the original Proto Tanna root \*im<sup>w</sup>a (minus the article) with the accretion of a locative prefix *i-*: thus \*i-im<sup>w</sup>a- (LOC-house/place-). Similarly, the form in AnejoŨ suggests the addition of a locative prefix *u-* onto the root *om<sup>w</sup>(a)*, thus Pre-AnejoŨ *u-om<sup>w</sup>a<sup>-5</sup>* (reconstruction, then, is \*(i,u)-i(u)m<sup>w</sup>a-).

Neither of these prefixes is currently productive. Both, however, are found as fused prefixes on a wide range of place names. In addition, *i-* in Tanna occurs in certain other locative forms (e.g. Lenakel *tehe* ‘sea’, *irhe* ‘seawards’;<sup>6</sup> *nelukə* ‘the middle of’, *ilukən* ‘between’). AnejoŨ *u* occurs as a locative marker in one unusual locative construction, and there is a probable cognate *u(n)-* in Sye and Ura which forms locative nouns from general nouns; this may derive from POC \*ua[tu] ‘towards hearer’ (Malcolm Ross, pers. comm.).

## 4.2 Nominalisation

Proto Southern Vanuatu underwent a process by which the POC common article \*na was accreted onto most common nouns (the exceptions need not concern us here), thus:

(11) POC	SYE	LENAHEL	ANEJOŨ	
*na lipon-	<i>nelve-</i>	<i>nelu-</i>	<i>nejhe-</i>	‘tooth’
*na Rum <sup>w</sup> aq	<i>nimo</i>	<i>nim<sup>w</sup>a</i>	<i>niom<sup>w</sup></i>	‘house’
*na qauR	<i>nau</i>	<i>nau</i>	<i>nau</i>	‘bamboo’

In this process of accretion, the \*a of the article was normally lost if the root began with \*Ca. The resulting initial *nC* cluster remained *nC* in underlying forms, though in surface forms in Tanna languages a schwa is inserted between the *n* and the consonant, while a prothetic *i*

<sup>5</sup> POC \*u is regularly reflected as AnejoŨ *o*. POC vowels were lost in absolute word-final position in AnejoŨ, as in \*Rum<sup>w</sup>aq > *n/iom<sup>w</sup>* ‘house’, \*kita > *e/yet* ‘see (intransitive)’, but not when the root-final vowel was followed by a suffix, as in \*Rum<sup>w</sup>aq > *um<sup>w</sup>a-* ‘PLACE possessive marker’, \*kita > *e/yta-i* ‘see (transitive)’.

<sup>6</sup> Both of these forms derive from POC \*tasik ‘sea’: \*t regularly becomes \*r, but when the POC article \*na fuses with initial \*t (see §4.2 below) the two coalesce as *t*.

occurs before the *n* in Anejoĩm (so the surface forms of the Anejoĩm words listed below are *inha-*, *inma* and *intal*):

(12)	POC	SYE	LENAKEL	ANEJOĨM	
	*na paqan-	<i>nva-</i>	<i>nəva-</i>	<i>nha-</i>	'thigh'
	*na maRi	<i>nmar</i>	<i>nəm</i>	<i>nma</i>	'breadfruit'
	*na talos	<i>ntal</i>	<i>nəte</i>	<i>ntal</i>	'taro'

The POc FOOD marker in Tanna and Anejoĩm has undergone this same accretion and loss of \*a:

(13)	POC	PSV	WHITESANDS	LENAKEL	KWAMERA	ANEJOĨM	
	*na ka-	*n-ya-	<i>nəŋ-</i>	<i>nək-</i>	<i>sa/na-</i>	<i>n ya-</i>	FOOD

The somewhat unusual reflexes of \*k (Whitesands *ŋ*, Kwamera  $\emptyset$ ) are illustrated in (14):

(14)	POC	WHITESANDS	LENAKEL	KWAMERA	
	*manuk	<i>menəŋ</i>	<i>menuk</i>	<i>menu</i>	'bird'
	*kayu	<i>nəŋi</i>	<i>nək</i>	<i>n/ei</i>	'tree'

The POc DRINK marker in Tanna (though not Anejoĩm) has also undergone the same accretion (though Proto Southern Oceanic seems to have irregularly changed \*m to \*m<sup>w</sup> in this morpheme):

(15)	POC	PSV	WHITESANDS	LENAKEL	KWAMERA	ANEJOĨM	
	*na ma-	*[n-]m <sup>w</sup> a-	<i>nəm<sup>w</sup>-</i>	<i>nəm<sup>w</sup>-</i>	<i>sa/nm<sup>w</sup>u-</i>	<i>{lum<sup>w</sup>a-}</i>	DRINK

It is probable that the Proto Tanna PLANT marker \*nai- (= \*n-ai-?), which marks a noun specifically as something which the possessor has planted, also shows this accretion. An obvious POc source is \*na kayu 'ART tree'. However, if this is the case, then the reflexes of the \*k in \*kayu in this morpheme are different from those given in (14).

The effect of this nominalisation has been to convert the possessive markers into directly possessed common nouns. Indeed, the place marker discussed in §4.1 can also be viewed in the same way, except that it is a directly possessed locative rather than a common noun. This development may be related to the development of possessive classifiers from nouns which has taken place in other Oceanic languages (and which was probably taking place in Proto Oceanic itself). The difference with the FOOD and DRINK markers is that here we have markers becoming nouns, rather than nouns becoming markers.

### 3.3 *lu*-accretion in Anejoĩm

The second syllable of the Anejoĩm DRINK marker *lum<sup>w</sup>a-* clearly reflects the POc marker \*ma-, and thus the Anejoĩm marker also appears to be historically bimorphemic but, unlike in the Tanna languages, it has not been prefixed with the article \*na.

Now, just as Proto Tanna created an additional possessive marker (\*n-ai- 'plant'), so Anejoĩm has also created one further marker—*liθa-* JUICE—as in (16):

## (16) ANEJOŃ

*neto*            *liθa-k*  
 sugarcane    POSS.JUICE-1SG  
 'my sugarcane (to suck the juice from)'

*neθeθ*    *liθa-i*            *nhalav*  
 breast    POSS.JUICE-CS    child  
 'the child's (mother's) breast (to suck the milk from)'

If we analyse *lum<sup>w</sup>a-* as containing an unknown morpheme *lu-* plus a reflex of POC \**ma-*, we could then suggest (i) that *liθa-* is also historically bimorphemic and (ii) that the initial syllable of *liθa-* may be related to the initial syllable of *lum<sup>w</sup>a-*.

There is some evidence to support assumption (i) AnejoŃ *θ* derives from POC \**s* (as well as from \**c*), and the second syllable of *liθa-* may derive from \**susu* 'suck, breast' (AnejoŃ *e/θeθ* 'suck', *n-e/θeθ*, *na/θe-* 'breast'), or from \**suRuq* 'juice' (AnejoŃ *ni/θi-* 'juice'). The vowels show variation—but this is almost a trademark of Southern Vanuatu languages!

There is also some evidence to suggest that assumption (ii) may be correct—that the first syllable of each of these markers may have the same origin, deriving from Pre-AnejoŃ \**lu*. Although the regular reflex of POC \**u* is *o*, there is a (weak) tendency for \**u* > *u* adjacent to *m<sup>w</sup>* and *p<sup>w</sup>*; and there is a strong tendency for POC \**u* to become *e* and occasionally *i* adjacent to *θ*. In addition to the reflexes of \**susu* and \**suRuq* given above, note also \**paluca* > *a/heleθ* 'paddle', \**qanusi* > *elw-aneθ* 'spit' (the first element from \**luaq* 'vomit') and \**nusi* > *niθ* 'octopus'.

Thus we appear to have two historically bimorphemic forms here, suggesting earlier \**lu-m<sup>w</sup>a-* DRINK and \**lu-sa-* JUICE. However, I am unable at this stage to determine what the origin of \**lu-* might have been, nor can I explain why it and not \**na* is prefixed to these two markers.

## 5 General possession

The POC GENERAL marker \**na-* (with its variant \**a-*) is widespread in northern and central Vanuatu, where the Southern Vanuatu family's closest relatives are spoken, although in some of these languages it has undergone a change to *no-* (and in some others to *ne-*). Unlike \**ka-* FOOD and \**ma-* DRINK, however, this marker has been completely lost in Southern Vanuatu. One possible explanation for this is that, as I have suggested elsewhere (Lynch 1996:106), GENERAL \**na-* may have been, or may have derived from, the common article \**na*. The loss of \**na-* GENERAL in Proto Southern Vanuatu may be associated with the loss of the article \**na* as an independent morpheme in these languages.

The AnejoŃ GENERAL marker is *u*;<sup>7</sup> alone among AnejoŃ possessive markers, *u* does not take the construct suffix when followed by a noun possessor:

<sup>7</sup> This marker has the allomorphs *uñā-* before *-k* 1SG, *uñu-* before *-m<sup>w</sup>* 2SG and *uwu-* ~ *owu-* before *-n* 3SG; *uwu* is occasionally used instead of *u* before a noun possessor. The nature of these forms, in comparison with *u* as used in other environments, suggests that they too may have been bimorphemic.



- (17) ANEJOŃ  
*niom<sup>w</sup>* *u*                      *ŋi?*                      *\*niom<sup>w</sup>* *u-i*                      *ŋi?*  
house POSS.GENERAL who                      house POSS.GENERAL-CS who  
‘whose house?’

This marker may well be related to the locative *u* discussed in §4.1, but appears to be formally unrelated to the GENERAL markers in any other Southern Vanuatu languages. I will not deal with AnejōŃ any further in this section.

The other languages all appear to have historically bimorphemic markers, and it seems that one of the morphemes involved may have been POc *\*sa-* (see §5.1.1 below). I say “may” because the sibilants have been rather more unstable in Southern Vanuatu languages than any other consonants: *s* and/or *h* in one or more Southern Vanuatu languages derive from *\*t* before a front vowel, *\*s*, *\*c* and *\*j*. Modern Sye is particularly notorious for the fluctuation between /s/ and /h/ (Crowley 1997), but there is a certain amount of synchronic fluctuation between these two phonemes in the Tanna languages. At the same time, I have had considerably more difficulty in reconstructing the Proto Southern Vanuatu sibilants and showing their development from Proto Oceanic than with any other class of consonants in these languages, although in general terms *\*s* > *s* ~ *h* in Sye, *s* in Kwamera and *h* in the other Tanna languages.

## 5.1 Tanna

The Tanna GENERAL possessive markers are all (at least) disyllabic, and appear to be historically bimorphemic. They all involve a certain amount of morphophonemic alternation, which is suggestive of the bimorphemic nature of these markers. For example:

- (i) the Lenakel marker *taha-* occurs in that form before singular pronouns and before nouns, but occurs as *tə* before nonsingular pronouns;
- (ii) in Kwamera, the form is *sava-* before third person pronouns and the construct suffix but *sa-* before other pronouns.

### 5.1.1 Northern Tanna

The forms in the three northern Tanna languages (North Tanna and Whitesands *raha-*, Lenakel *taha-*) are cognate. The second syllable would derive from earlier *\*sa*, and I have suggested that *\*sa-* is in fact one of the GENERAL markers that can be reconstructed for Proto Oceanic. Cognate forms are found in Western Oceanic languages like Takia, Torau and Mono-Alu, and in Central-Eastern Oceanic languages like Atchin, Port Sandwich, Lewo and possibly Paamese and Proto Polynesian; see Lynch (1996:107–108) for further discussion.

The first syllable of the northern Tanna markers shows a phoneme correspondence *r:r:t*, which apparently also derives from POc *\*s*, but in the environment of a *\*q* in an adjacent syllable, with the intervening vowel being lost.<sup>8</sup> Compare:

<sup>8</sup> The first example in (18) also shows another change conditioned by *\*q*: POc *\*n* > *ŋ* in the environment of *\*q* in an adjacent syllable (cf. *\*tinaqe-* > Lenakel *nəŋqaa-* ‘intestines’, *\*qanusi* > Lenakel *aŋh* ‘spit’, and so on).

(18) POC	N. TANNA	WHITESANDS	LENAKEL	
*qasan	ne/rŋ-	ne/rŋ-	ne/tŋ-	'name'
*saqat	a/raat	ɔ/ra	taat	'bad'
*leqos 'look at'	er/anən		eit/anən	'look at reflection'

This suggests, therefore, that the northern Tanna GENERAL markers derive from something like \*sVqa-sa or \*qVsa-sa. I have no idea where the \*sVqa-/qVsa- form may have come from.

### 5.1.2 Southern Tanna

The form \*sa- is reflected in the first syllable of the Kwamera form *sava-*. The second syllable would derive from an earlier \*pa-, but I have no idea what its origin might be.

In Southwest Tanna, the Nəvhaal dialect has *kapa-* ~ *kafa-*, while the Nəvai dialect has *kapaha-*. There are many cases of \*pVh or \*hVp coalescing as *f* in the Tanna languages, and I suggest that the underlying Southwest Tanna form is *kapaha-*, with the final syllable regularly deriving from \*sa-. Now Kwamera has another, relatively rare, passive marker (in addition to the more common *ia*, *ira-*) that has the form *kəfu-* before the first person singular suffix and *kə* elsewhere. This appears to be phonologically related to the Southwest Tanna form, and suggests an earlier \*(k,g)aba-sa-. Again, the origin of \*(k,g)aba- is not known.

## 5.2 Erromango

Recent developments in the possessive-marking system of the Erromangan languages are discussed by Crowley (this volume). The system in (19) was probably the original one in Sye, with the system Crowley describes based on (*h*)*en-* + focal pronouns a later development, possibly based on the third person marker *ihe-* + *-n* (third person singular or construct suffix).

### (19) SYE

*horo-* ~ *hore-* ~ *hor-* before non-third person pronouns

*ihe-* before 3SG, 3PL and construct suffix *-n*

Sye *h* derives from POC \*s. The *hor(V)-* form may well derive from earlier \*sa-rV-, while the *ihe-* form may derive from \*i-sa-.

Ura seems to have had the marker *ar(V)-*, and Utaha the cognate marker *et-* or *etV-*. Although Ura *r* corresponds with Sye *r* (< POC \*r and \*R), it also corresponds with Sye and Utaha \*t (< POC \*t). Given the close relationship between Ura and Utaha and the similarity of the forms in both languages, I suggest that both the Ura and Utaha forms derived from an earlier \*at(V)-. Now there are also POC GENERAL markers \*a- and \*ta-, and it is **possible** that this form was originally \*a-ta-.

There is, however, another possibility. Sye initial *s* (or *h*) corresponds both with Ura *s* and  $\emptyset$ , as illustrated in (20).

(20) a.	SYE <i>s-</i>	URA <i>s-</i>		cf. POC
	<i>si-</i>	<i>si/n</i>	'excrement'	*taqe-
	<i>selai</i>	<i>selai</i>	'shine light on'	
	<i>sorvat</i>	<i>sorvat</i>	'remove stones from fire'	
	<i>suju</i>	<i>suju</i>	'kiss'	
	<i>sanwis</i>	<i>sanwis</i>	'wild boar'	
b.	SYE <i>s-</i>	URA $\emptyset$ -		cf. POC
	<i>sei</i>	<i>ai</i>	'to spear'	*sua ?
	<i>sat</i>	<i>arw-at</i>	'bad'	*saqat
	<i>soki</i>	<i>e yi</i>	'climb up, copulate'	*sake
	<i>savel</i>	<i>afel</i>	'whistle'	
	<i>semsi</i>	<i>amsi</i>	'choose'	

The *s*: $\emptyset$  correspondence seems to be found only in verbs and adjectives. However, it is possible that the Ura form *ar(V)* could derive from \**sa-ta-*, and that the first element is thus cognate with the Sye form.

### 5.3 Summary

To summarise the discussion, the earlier forms of the markers seem to have been as listed in (21). Forms in bold possibly reflect Proto Oceanic reconstructed possessive markers.

(21) ERROMANGO	
SYE	<b>*sa-ra-</b> , <b>*i-sa-</b>
URA, UTAHA	<b>*a-ta-</b> or <b>*sa-ta-</b>
TANNA	
NORTHERN	*qVsa- <b>sa-</b> or *sVqa- <b>sa-</b>
S.W. TANNA	*(k,g)aba- <b>sa-</b>
KWAMERA	<b>*sa-pa-</b> , (*(k,g)aba- <b>sa-</b> ?)
ANEITYUM	
ANEJOM	*u[ ]-

The following comments can be made:

- (i) All languages except Anejoṃ have a reflex of \**sa-* as one of the morphemes involved in the GENERAL marker. The Anejoṃ form may have been simply \**u-*, or it may have had some additional phonemic material.
- (ii) The other morphemes involved do not appear to be cognate.
- (iii) It is possible that the disyllabic forms (northern Tanna \*qVsa-/sVqa- and southern Tanna \*(k,g)aba-) are themselves bimorphemic.

## 6 Kwamera *sa*-accretion

I return now to a topic I left a little while ago: the fact that a number of the Kwamera markers begin with *sa*. First, the POC FOOD and DRINK forms have accreted \**na-*, then \**sa-*, as follows:

(22)	PSV	PRE-KWAMERA	KWAMERA	
	*na-ka-	*sa-na-ka-	sana-	FOOD
	*na-m <sup>w</sup> a-	*sa-na-m <sup>w</sup> a-	sanm <sup>w</sup> u-	DRINK

The Kwamera PLANT marker *sapwas-* does not derive from Proto Tanna \*n-ai-. Rather, it probably derives from the verb *apwe* 'to plant', and may thus derive from an earlier \*sa-(a)(bw,bu)V-sa-. The GENERAL marker, as noted in the previous section, derives from a Pre-Kwamera form \*sa-pa-.

What seems to have happened is this: Proto Southern Vanuatu, as I mentioned earlier, nominalised most possessive markers, and thus converted them into inalienable nouns. Pre-Kwamera, but no other Southern Vanuatu language, seems to have reasserted their function as possessive markers by prefixing the marker *sa-* to them. Most of the Kwamera possessive markers, therefore, are historically trimorphemic, as in (23):

(23)	sanm <sup>w</sup> u-	< *sa-na-m <sup>w</sup> a-
	POSS.DRINK-	POSS-ART-POSS.DRINK-

## 7 Discussion

Proto Southern Vanuatu nominalised the possessive markers used in indirect constructions, converting them into inalienable (common or locative) nouns. The two Proto Oceanic systems—possessive markers proper and nouns used as possessive classifiers—seem to have collapsed into a single system in Proto Southern Vanuatu. Reflexes of \*ka- FOOD and \*ma- DRINK (and \*sa- GENERAL?) show retention of possessive markers, while the PLACE, PLANT and JUICE markers illustrate the development of possessive classifiers.

This in itself is not a particularly unusual development. Possessive markers function grammatically like directly possessed nouns, in that they take possessive pronominal suffixes directly. Since many such nouns would have been preceded by an article in Pre-Proto Southern Vanuatu (either \*na or some other article, like \*e 'kin'), there would have been paradigmatic pressure on the possessive markers to be similarly marked. Note also that possessive marker + pronoun can stand alone in the Southern Vanuatu languages as the sole constituent of a noun phrase or even of a clause:

(24)	LENAKEL	
	<i>Pehe r-əm-os</i>	<i>nək-n?</i>
	who 3SG-PAST-take	POSS.FOOD-3SG
	'Who took his (food)?'	

(25)	ANEJOM
	<i>Lum<sup>w</sup>a-k!</i>
	POSS.DRINK-1SG
	'It's mine!' (often shouted when a mango is heard falling from a tree)

So the conversion of possessive markers to inalienable nouns can be seen as a fairly natural development. What does need explanation, however, is why the GENERAL marker shows no evidence of the accretion of the article \*na; though the GENERAL markers in all Southern Vanuatu languages are also (at least) bimorphemic, there is no evidence that the first of these morphemes is an article.

Kwamera is unique among Southern Vanuatu languages in having taken the nominalisation process one step further, by prefixing a possessive marker to the nominalised forms. The widespread POc GENERAL marker \*na- was completely lost in Southern Vanuatu, and was apparently replaced by the less widespread POc marker \*sa- which, however, always seems to have occurred in association with some other morpheme. What seems to have happened in Kwamera is (i) the standard Proto Southern Vanuatu nominalisation, converting possessive markers to directly possessed nouns, and then (ii) the prefixation of sa-, apparently converting a noun back into a marker again.

## References

- Crowley, Terry, 1997, The voiceless fricatives [s] and [h] in Erromangan: one phoneme, two, or one and a bit? Paper presented at the Third International Conference on Oceanic Linguistics, Hamilton, New Zealand.
- 1998, *An Erromangan (Sye) grammar*. Oceanic Linguistics Special Publication No. 26. Honolulu: University of Hawai'i Press.
- Lynch, John, 1992, "For my part ...": the grammar and semantics of part possession in the languages of Tanna. *Australian Journal of Linguistics* 12:249–270.
- 1996, Proto Oceanic possessive-marking. In John Lynch and Fa'afo Pat, eds *Oceanic studies: Proceedings of the First International Conference on Oceanic Linguistics*, 93–110. Canberra: Pacific Linguistics.
- Lynch, John, Malcolm Ross and Terry Crowley, forthcoming, *The Oceanic languages*. London: Curzon Press.



# *Personal nominal words in Indonesian: an anomaly in morphological classification*

---

WARUNO MAHDI

Conventional morphological classification traditionally divides the nominal words into nouns and pronouns, but Indonesian morphological patterns argue for a cross-cutting distinction between personal nominals (proper names, personal ‘pronouns’ and ‘pronominalised’ words) and nonpersonal nominals (nouns, demonstrative and interrogative ‘pronouns’) on purely grammatical considerations. This paper inspects the grammatical categories of the personal nominals that distinguish these from the nonpersonal nominals already treated elsewhere (Mahdi 1993:183–200). Particular attention is dedicated to ergativity as a feature of the personal nominals not shared by nonpersonal nominals. This reveals a further anomaly shared, however, with some other languages (for example Tagalog), in that Indonesian combines features of ergative and accusative languages. Other distinctive features of the personal nominals discussed here are the categories of person and grammatically relevant abridgment, the vocative, the expression of relative social rank, and that the nonplural number is singular (rather than nonspecific as in the nonpersonal nominals).

## 1 Introduction<sup>1</sup>

Proper names traditionally figure as a subclass of nouns, this in turn being morphologically distinct from a class of pronouns that encompasses personal, demonstrative, relative and interrogative pronouns. An earlier publication (Mahdi 1993) pointed out that Indonesian Malay offers an alternative grouping that seems to be at least anomalous, insofar as I am not aware of the same situation having been reported elsewhere. The main division in the

---

<sup>1</sup> Knowing Byron, I am probably only one in a very long list of persons who in one way or the other owe him many thanks. Paradoxes in morphology seem to be one of his pet subjects, and so I hope that the subject of this paper will pass for a little thank you, even if an anomaly does not quite reach the subversive effect of a full-fledged paradox on conventional wisdom. I am indebted to the Fritz Haber Institute, particularly the Department of Physical Chemistry headed by Gerhardt Ertl, for the generous permission to use Institute facilities for my private linguistic studies.

nominals appears to be not between nouns and pronouns, but between a class that encompasses nouns, nonpersonal (e.g. geographical) proper names and nonpersonal (e.g. demonstrative, interrogative) pronouns on the one hand, and one that includes personal proper names and pronouns on the other. Strictly speaking, therefore, one cannot actually speak of 'pronouns' as a word class in Indonesian.

The class of nonpersonal nominals was discussed in some detail in Mahdi (1993:183–200), and the present paper seeks to analyse the personal nominals. Three subclasses of the latter will be distinguished: personal proper names, personal pro-names<sup>2</sup> (personal pronouns), and relational pro-names (these are personalised kinship terms and honorific titles functioning as personal pronouns). Personal articles, being similar to an article mode in the paradigm of relational pre-names, will also be discussed.

The internal grouping within the two nominal classes can thus be schematised as follows (divergent terms used in Mahdi 1993 are added in parentheses):

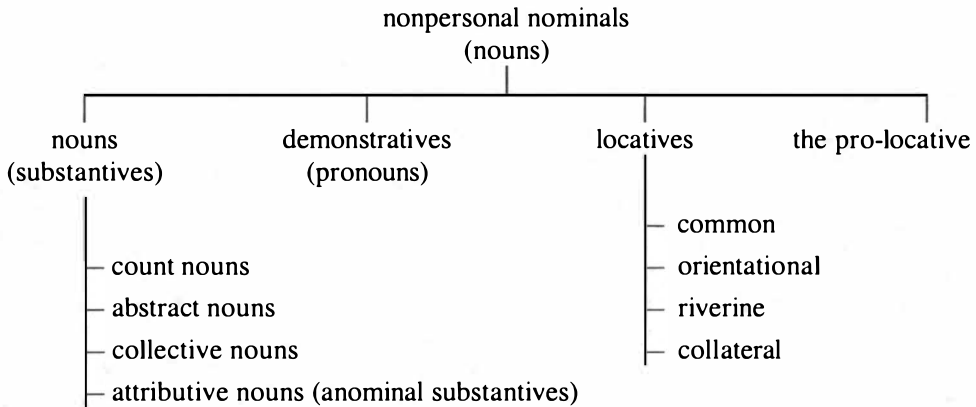


Figure 1: Internal grouping of nonpersonal nominals

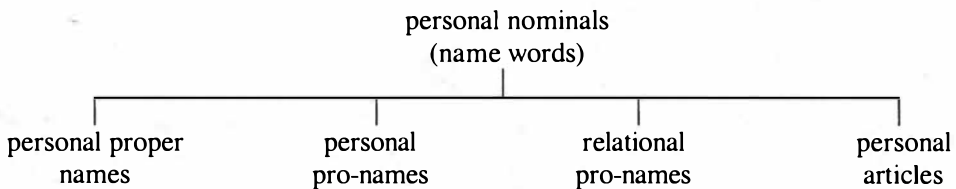


Figure 2: Internal grouping of personal nominals

<sup>2</sup> It seems to be a particularity of Indonesian grammar that not only does it have 'substitute words' for nouns (as conventional grammar leads us to expect) in the form of pronouns, but also features these for several word classes (including some not indicated here). The name of each class of 'substitutes' is formed here by prefixing *pro-* to the name of the class to be replaced, except that 'demonstratives' is used instead of 'pronouns' to avoid confusion with 'pronouns', which, as already indicated, expresses a concept that is not applicable to Indonesian grammar.



By contrast to the Indonesian nonpersonal nominal, the personal nominal cannot function as qualitative attributive. In the nonpersonal nominals, the nonplural number is not the singular, but non-specific, and indication of actual quantity is provided by an attribute or in the context. The use of plural form lies in either specifying the quantity in absence of such attribution or context, or indicating variety (or both). In the personal nominals, on the other hand, the nonplural is the singular and expresses the precise quantity of oneness. Unlike the nonpersonal nominals, furthermore, the personal nominals distinguish grammatical case (ergative versus nonergative, in some subclasses also a vocative) and abridgement (grammatically distinct long versus short forms). Finally, the personal nominal expresses grammatical person (at least in some of its forms); that is, it refers to first, second or third person, or a specific combination of these.

A similar division into personal and nonpersonal nominals will perhaps be found in other languages, particularly in the closer circle of other Austronesian languages. Indeed, the phenomenon under discussion in Indonesian probably reflects tendencies already inherent in early stages of development of the language family or some of its subdivisions. In an early discussion of nouns in Jabêm, Dempwolff (1939:19) provides evidence of distinctive grammatical properties indicating a fundamental dichotomy between two major classes, those denoting individuals (*Einzelwesen*) and those denoting types (*Gattungen*). According to Dempwolff, the distinction often equates to one between person (*Person*) and thing (*Sache*).

In a much more recent summary for Oceanic languages, Lynch (1998:100) begins his grammatical overview thus: "I use the term 'pronoun' fairly loosely. Oceanic languages generally have only one set of free pronouns, but they also have one or more sets of pronominal forms that are more or less bound to nouns, verbs, or other morphemes. While only the free forms might qualify as pronouns under a strict definition, I discuss the other forms here as well."

However, the distinction in Indonesian differs from situations reported so far for related languages. On the one hand, it is not semantic in nature, but purely grammatical. Thus the nouns (included among the nonpersonal nominals) include people-denoting words such as *bèsan*<sup>3</sup> 'parent of son/daughter-in-law', *ipar* 'sibling-in-law', *keponakan* 'niece/nephew', *kerabat* 'relative', *laki-laki* 'man (male person)', *manusia* 'human being', *perempuan* 'woman', and *tetangga* 'neighbour'. The denotation of personal and nonpersonal nominals refers to the relevance of grammatical person to the meaning or use of the word (whether or not reference to the first, second or third person, or some combination of these, is implied).

On the other hand, personal nominals corresponding in general function and meaning to the (personal) pronouns of conventional grammars share their principal distinctive features with personal proper names and relational pro-names. The latter etymologically derive from certain people-denoting nouns (being nonpersonal nominals), particularly kinship terms and titles, having acquired the characteristic distinctive grammatical features of personal nominals

<sup>3</sup> Indonesian data in this work are given with the following deviations from the currently valid standard spelling: a clitic is joined to, or separated from, the prosodic base word by an equal sign, e.g. *dia=pun* 'and he [thereupon]', *dia=nya* 'he, him there' (standard spelling has *dia pun*, *dianya*); the mid central, mid front and low front unrounded vowel phonemes are spelled *e*, *é* and *è* respectively (as in French), whereas standard Indonesian spelling has *e* for all three, and standard dictionary transcriptions have *e* for the first, and *é* for the other two. In a closed syllable, the pronunciation of *é* approaches that of English *i* in *bit*, *bid*, *bin*. When dialectal divergence in the realisation either as *é* or as *è* exists, the one that seems less regionally specific is assumed. Finally, the mid back rounded vowel is spelled *ó* to distinguish it from the low back rounded *o*. The standard spelling has *o* in both instances (also in standard dictionary transcriptions).

upon being 'pronominalised'. The people-denoting nouns involved do not include any of the ones listed in the previous paragraph.

But whether or not the same or a closely similar situation should become apparent in other languages, it provides a grouping of word classes that is irreconcilable with that to which one is traditionally accustomed, at least in that it makes the conventional notion of a 'pronoun' inapplicable, and excludes personal proper names—though not topological and astronomical ones—from the nouns.

## 2 Personal pro-names

### 2.1 Person, number and social ranking

Although it would seem logical to first begin with the proper names, and only after that to proceed with the pro-names that serve as proper-name 'substitutes', the reverse order will be followed here, because the class of personal pro-names features the most elaborate paradigm of forms. It is therefore simpler to discuss the grammatical categories of the personal nominals in this subclass.

The personal pro-names are the semantic correspondents of what are conventionally referred to as personal pronouns. Like these, they distinguish number and person. But here, the category of number is practically a semantic one and has no grammatical consequence. Some personal pro-names have a single referent, others more than one.

In the personal pro-names, the category of person likewise has no grammatical consequence, but this is not entirely the case in the other subclasses of personal nominals, where there may be a kind of relative concord in respective use by speech partners in a dialogue (I will return to this below). Indonesian is one of the languages that distinguish an 'indefinite person' (cf. French *on*, German *man*), actually an unspecified or abstract third person.

Another category in the Indonesian personal pro-names of marginal grammatical consequence, though widespread in languages of Southeast Asia, is not usually included among the features traditionally ascribed to the conventional personal pronoun. It is the category of relative social ranking that appears in the paradigm of personal pro-names as a distinction between formal and informal varieties.

The distribution of personal pro-names according to these categories is laid out in the following table<sup>4</sup>, in which only the long forms are included. The short forms will be discussed separately below.

<sup>4</sup> The table does not include colloquialisms and dialectisms. Deserving mention among these are: the Betawi (Jakartan Malay) *gué* or *gua* '1', *elu* '2'; Ambon Malay *bêta* '1', *osé* '2'; Arab Malay *ana* '1', *énté* '2'; and the use of *kita* (also *kité*) for '1' in a formal but congenial Betawi environment. Forms identified above as being Betawi also pertain to youth and colloquial speech styles.

Table 1: Personal pro-names

SINGLE REFERENT			MULTIPLE REFERENT	
INFORMAL	FORMAL			
<i>aku, saya</i>	<i>kami</i>	'1' <sup>5</sup>	<i>kami</i>	'1+11 / 11+33'
<i>engkau, kamu</i>	<i>anda</i>	'2'	<i>kalian</i>	'2+22 / 22+33'
<i>dia</i>	<i>beliau</i>	'3'	<i>merèka</i>	'3+33 / 33+33'
			<i>kita</i>	'11+22 / 11+22+33'
	<i>orang</i>	'00'		

The words *kami* (with multiple referent) and *kita* are generally glossed as 'we (exclusive)' and 'we (inclusive)', respectively. The pro-name of the indefinite person (*orang*), like the count-word for persons (*orang*), derives etymologically from the noun *orang* 'person'. It has the particularity of being nonspecific in number of referents. When *kami* is used as a formal, single-referent pro-name of the first person, on the other hand, it explicitly refers to one person (it is as though there were two words *kami*, being metonyms).

The single-referent personal pro-name also expresses the relative social ranks of the speaker/writer and the listener/reader or a third person referent. The items listed under 'informal' are used in speech with or about a person of approximately equal or lower rank, or in informal speech, particularly between peers. Otherwise, the use of these pro-names can have an imposing, sometimes even pejorative effect. Where more than one entry is given under the same rubric, the one to the right is somewhat less imposing.

In contrast with the situation in many European languages, the use of *kami* for '1' is more modest than that of *saya* or *aku*. Nevertheless, in translations of European texts the 'royal plural' is generally rendered into Indonesian with *kami*, probably because it at least conveys the formality of the occasion, if not accurately expressing the superior rank of the referent.

Another semantic particularity needs to be noted for pro-names of the third person singular *dia*, *beliau* and plural *merèka*: they only refer to persons. Hence, *dia* does not translate English *it* (except when referring to persons as in *it, the child/baby*).

Items in the right column are not the plural grammatical forms of the corresponding items in the left column (where there is one). The pairs in each row are lexically distinct, such as English *person, people*. The personal pro-names do however have a reduplicated form resembling that of the plural of the nonpersonal nominals, which one could call the emphatic plural, as for *saya* '1' and *merèka* '3+33'. Only *orang*, the indefinite-person pro-name, does not seem to have a reduplicated form.

*Saya-saya lagi=lah yang disalahkan.*  
'And it's **me** again who gets the blame.'

<sup>5</sup> The following abbreviations are used: 0 – indefinite person, 1 – first person, 2 – second person, 3 – third person, 00/11/22/33 – one or more individuals of the respective person, AI – intransitive agent, AT – transitive agent, AUX – auxiliary, CIRC – circumstantial complement, O – accusative object, P – patient, S – nominative subject, V – verb; / indicates alternative meanings.

*Jangan tanya saya, merèka-merèka itu=lah yang lihat.*

'Don't ask me, **they** are the ones who saw it.'

As the multiple-referent ones have a reduplicated form, and the single-referent first person pro-names refer to one individual, one might question the rationality of calling these forms the emphatic plural. Nevertheless, the reduplicated form of the single-referent second and third person pro-names can refer to several individuals. Compare for *kamu* '2':

*Mengenai itu, cukup kalau kamu-kamu ini saja yang mengetahui.*

'As for that, it suffices that you people alone know of it.'

In other words, actual 'plurality' is only expressed in the emphatic plural form of some of the personal pro-names. However, as we shall see below, the reduplicated form of the relational pro-names all express plurality, and do so without the emphasis of this form of the personal pro-names. Furthermore, a plural formed by reduplication is the only grammatical feature the pro-names have in common with the nonpersonal nominals in general, and with the demonstratives in particular. The implication of this circumstance for typological comparisons with languages having a canonical class of pronouns (personal and demonstrative/interrogative in one class) is, however, strongly diminished by the totally divergent structural position of the plural in the Indonesian personal pro-names compared to that in canonical personal pronouns.

Finally, there is another usage involving single-referent personal pro-names to express the respective plural denotation that is mainly restricted to colloquial speech. This seems to be a feature borrowed from Ambonese Malay, which is in turn apparently an early calque from colloquial Dutch. The plural forms are derived by suffixing *orang* 'person' (calquing Dutch *lui* 'people, folk') to *kita* (as colloquial for '1'), *kamu* and *dia*,<sup>6</sup> as in the following:

*Kamu-orang lagi mau ke=mana?*

'Where are you-people meaning to go to?'

## 2.2 The ergative

The most important grammatical category in the morphology of the personal nominals, by which they contrast with the nonpersonal nominals, is that of case, which is typically represented by the opposition between ergative (Mahdi 1993:200) and nonergative case forms. The ergative form is a proclitic, pronounced together with the following word as one would pronounce a compound word. Otherwise, the opposition is unmarked, except in a subclass of the personal pro-names that will be considered separately below. Apart from that, there are syntactical particularities that are common for all the personal nominals.

However, the identification of the ergative as such in Indonesian is not unproblematic. Although the term has been used in a variety of meanings, one has with some exceptions tended to apply the definition given by Dixon (1979:61–62). The author classifies a language as ergative "if intransitive subject is treated in the same manner as transitive object, and differently from transitive subject" (see also Dixon 1972:128, 1994:1; Du Bois 1987:807;

<sup>6</sup> In Ambonese Malay itself, these derivations have undergone further simplification, e.g. *dorang* or even *dong* 'they' for *dia-orang*.

Manning 1996:3). Where nominals are distinctively case-marked, the differently treated transitive subject is in the ergative case.<sup>7</sup>

One could provide the following formal grounds for qualifying Indonesian as an ergative language. In current conventional terminology, the grammatical form assumed by AT in (2a–b) is the ergative case, that assumed by AI in (1) and P in (2a–b) the absolutive case (Dixon 1994:1):

- (1) *Aku mengajar.* 'I teach.'  
 AI V
- (2a) *Engkau ku= ajari.* 'You are taught by me.' (lit. 'you, I teach')  
 P AT V
- (2b) *Aku kau= ajari.* 'I am taught by you.' (lit. 'I, you teach')  
 P AT V

However, although these patterns formally qualify Indonesian as an ergative language, they only tell half the truth with regard to the morphology and syntax of the language. Indonesian can probably be included among those languages featuring absolutive–ergative as well as nominative–accusative structures. The parallel coexistence of the two structures in Malay seems to have been noted by Hopper (1983) in a work I unfortunately have not been able to access (cited here from Du Bois 1987). The nominative–accusative structure in Indonesian can be demonstrated by the following:

- (3a) *Aku mengajari =mu.* 'I teach you.'  
 AT/S V P/O
- (4a) *Engkau diajari oleh =ku.* 'You are taught by me.'  
 P/S V by AT/O

Traditionally, the construction exemplified in (3a) was regarded as an expansion of (1), both being subsumed under active voice. In descriptions of School Malay (the prewar standard), construction (2a–b) passed for the complementary passive voice for first and second person singular agent. This tradition was continued in the postwar period in standard Indonesian (see Moeliono & Dardjowidjojo 1988:94).

Construction (4a) for a first or second person singular agent, on the other hand, appeared in the unofficial Malay of the prewar press and of political and other public debates as an alternative passive voice (apparently under influence of Sundanese, see Zain 1940:21). For third person singular and plural, first and second person plural, and for nouns, this construction was already standard for passive voice in prewar School Malay; compare (3b) and (4b–c) (the terms in square brackets are optional).<sup>8</sup> In other words, the paradigm that

<sup>7</sup> The more widely used, but to my mind less consistent, abbreviation introduced by Dixon (1972:128, 1979:61) has S (from 'subject') for my AI, A (from 'agent') for my AT, and O for P (the latter also used in place of O by Comrie 1978:332). After all, Dixon's S is also an agent, while the logical counterpart to agent (Dixon's A) is patient (P), not object (O). But I find the received terminology generally misleading, because in a nominative–accusative structure, a subject is agent in the active voice, but patient in the passive, and vice versa for object. The presently proposed terminology seems less eclectic, reserving 'subject' and 'object' for their more traditional use for the respective syntactic modes in nominative–accusative sentences.

<sup>8</sup> Constructions of the type (4b), but with a postpositioned patient referent, were classified as ergative in an early publication by Herrfurth (1983:190): *Dibeli=nya rumah* 'He [thereupon] bought a house' (lit. 'bought [by] him [a] house'). But the author operated with a different definition of 'ergative', and seems to

passed for active/passive voice in School Malay and early standard Indonesian actually combined two different structures: a formally absolutive–ergative for first and second person singular pronominal agent, and a nominative–accusative for other pronominal and nominal agents.

- (3b) *Ia mengajari =ku.* '(S)he teaches me.'  
 AT/S V P/O
- (4b) *Aku diajari [olèh] =nya.* 'I am taught by him/her.'  
 P/S V by AT/O
- (4c) *Ia diajari [olèh] guru.* '(S)he is taught by the teacher.'  
 P/S V by AT/O

Incorporation of formerly nonstandard features of the language into modern usage has lifted the constraint that restricted use of the respective constructions to particular grammatical person and number. Thus, as already noted by McDonald and Dardjowidjojo (1967:235), one can nowadays observe personal pro-names of any person or number as agents in constructions (2a–b), as in (2c–d). The still remaining constraint is that only a personal nominal can occupy the position of **AT** in this construction. It is thus a feature of the personal nominals that distinguishes them from the nonpersonal nominals.

- (2c) *Ia kami=ajari.* '(S)he is taught by us.'
- (2d) *Kami ia=ajari.* 'We are taught by him/her.'

A further development is the use of the long form in place of the short form previously prescribed in prewar standard School Malay for the agent in (2a–b) and the active-voice patient in (3a–b) and passive-voice agent in (4a–b). Thus, for (2a–b) one can nowadays also have (2a'–(2b')):

- (2a') *Engkau aku=ajari.* 'You are taught by me.'
- (2b') *Aku engkau=ajari.* 'I am taught by you.'

As a result of all this, Indonesian now appears to have two parallel 'passives': that exemplified in (2a–b–c–d, 2a'–b'), and that in (4a–b–c); see Chung (1976) and Verhaar (1978). At least on the phonological plane, the different treatment of **AT** in (2a), compared to that of **AI** in (1) and **P** in (2b), shows up only in the proclisis of **AT** (implying also the suppression of word stress). However, the short form *ku* can still be used for **At** in (2a), whereas it cannot be used for **AI** in (1) and for **P** in (2b) (I will return to this below).

Significant, in my opinion, is that the grammatical meaning expressed by **AT** in (2a) is analogical to that in (4a).<sup>9</sup> In both constructions we have topicalisation of the patient, and a nontopicalised agent. It is this circumstance that led to both constructions being qualified as passive voice even in recent mainstream descriptions. Thus, constructions (2a–b) and

---

have mistaken the agent referent for a possessive attribute. (That it is not possessive becomes evident upon insertion of the optional instrumental preposition *olèh* 'by' between it and the preceding verb.)

<sup>9</sup> In Indonesian one reverts to a passivelike construction (i.e. one with topicalised patient) more readily than in English. One should not therefore be surprised that a nonliteral English translation correctly conveying the original meaning will often need the active voice where the original is structurally passive or ergative. It is even quite common to use the passive (*Diminum=lah tèh itu*) or ergative (*Kau=minum=lah tèh itu*) in an imperative sentence ('Do drink the tea').

(4a–b–c) are referred to respectively as “D passive” and “*di*-D passive” (D stands for verb base) in the standard grammar of Moeliono and Dardjowidjojo (1988:94), and as “passive type two” and “passive type one” in the reference grammar of Sneddon (1996:247–249). With regard to the first construction, classified here as ergative, Verhaar (1978:11–14) distinguishes between a “pronominal passive” when the agent is *ku=* or *kau=* (as in 2a–b, in which case the agent and verb are spelled as one in the standard spelling), and a “zero passive” otherwise (as in 2c–d, 2a’–b’, in which the two terms are spelled disjunctively), referring separately to construction (4a–b–c) as a “*di*-passive”.

Cartier (1979) seems to have been the first to have recognised that the (2a–b) alternative passive voice is actually ergative.<sup>10</sup> However, she identified the ergative sentence construction in Indonesian with the use of the ‘de-voiced’ (voice-neutral) transitive form of the verb. But it has been indicated elsewhere (Mahdi 1993:204) that there is also another voice-neutral verb form in Indonesian that does not require that the agent be in the ergative, for example:

*Peluru kena pada sasaran.* ‘The bullet hits the target.’  
 AT V at P

*Sasaran kena oleh peluru.* ‘The target is hit by the bullet.’  
 P V by AT

Cartier’s suggestion that the verb form involved in construction (2a–b) is voice-neutral as well nevertheless serves as further support for the classification of the corresponding construction as ergative.

Two circumstances can make it difficult to distinguish an ergative sentence from an active one in written language, because clitisation is not indicated in the spelling for polysyllabic personal pro-names (in deliberate and unhurried speech, the cliticisation is audible). One is that some verbs are transitive without distinctive marking, and the other is that Indonesian allows placing the grammatical object in nominative-accusative sentences—the patient in active voice (5a), the agent with instrumental preposition in passive voice (5b)—at the beginning of the sentence. An important distinguishing rule is that auxiliaries and adverbs are placed between nominative subject and verb, but before the ergative agent-verb group (so for a nominative–accusative NP it is NP–Aux–V, for an ergative it is Aux–NP–V). The following examples include *sudah* ‘already’ as adverb to the verb and *habis* ‘finished’ as circumstantial complement. Thus, besides clitisation as the only remaining phonological distinction, there is a difference in the syntactical behaviour of the personal pro-name in the nominative as in (5a), and in the ergative as in (6).

(5a) *Nasi itu kami sudah makan habis.*  
 P/O AT/S AUX V CIRC  
 ‘The rice we have eaten up.’

(5b) *Oleh kami nasi itu sudah dimakan habis.*  
 by AT/O P/S AUX V CIRC  
 ‘By us the rice has been eaten up.’

<sup>10</sup> Ergativity as a feature of Austronesian languages has been known for a long time for Tagalog in the Philippines and for Polynesian isolects. More recent studies suggest that it is not unusual in languages of Western Indonesia either. Beside having apparently already been indicated for Malay (Hopper 1983, cited in Du Bois 1987), ergativity has been reported for Sasak by Austin (1996) and for Balinese by Wechsler and Arka (1998).



- (6) *Nasi itu sudah kami= makan habis.*  
 P            AUX    AT            V            CIRC  
 'The rice has been eaten up by us.'

In principle, Indonesian does not forbid insertion of a word between a clitic and the prosodic base word, as the insertion of the optional *oleh* before *=nya* in (4b) demonstrates (in the absence of the instrumental preposition, the enclitic leans on the verb). A similar insertion after an ergative proclitic involves formation of a pro-name group that will be discussed separately below. The following preview example demonstrates one of its peculiarities—the proclitisation of the entire group in the ergative:

*Murid yang ku=ini=ajari mendapat angka-angka baik semua.*  
 'Pupils that I here taught all got good marks.'  
 (lit. '... were taught by this me ...')

This suggests that proclitisation of the ergative agent referent is probably not a morphological feature of the ergative case, but a syntactic feature of the absolutive–ergative sentence. However, as it is the only phonological feature by which one can discern the ergative from other case forms, and the ergative does not appear in any syntactic environment that does not require proclitisation, it is a useful distinguishing feature. For practical reasons, I will therefore consider it a mark of the ergative, within the personal nominals to be discussed.

In the preceding examples, the grammatical form of the personal pro-name appearing in the absolutive as **AI** in (1) and as **P** in (2a–b) is identical with that appearing in the nominative as **AT** in (3a–b) and as **P** in (4a–b). I will therefore refer to it as the nominative–absolutive case form. For pro-names that do not distinguish an oblique-genitive case, the nominative–absolutive case form also figures in the functions of direct and indirect object, and of possessive attribute. There is no grammatically distinguishable accusative case in Indonesian.

### 2.3 Abridgement

Personal nominals in Indonesian typically have a long and a short form that are not identical in their syntactic properties. This is another feature that contrasts them with nonpersonal nominals. However, not all personal pro-names have short forms. They can thus be divided into abridgeable personal pro-names (of which there are only four: *aku* '1', *kamu* '2', *engkau* '2' and *dia* '3') and unabridgeable ones (encompassing all the rest).

One particularity of the short forms of the personal pro-names is that they distinguish three cases instead of only two. For the short forms, there is an additional oblique-genitive case that the short pro-name adopts when serving as a complement or possessive attribute. The short ergative forms, like the long ergative forms, are all proclitics. The oblique-genitive (short) forms are all enclitics. The following table shows the morphological paradigm of the four abridgeable personal pro-names. In dialects not having the oblique-genitive *=kau*, the short forms for the second person singular thus make up a suppletive paradigm.



**Table 2:** Abridgeable personal pro-names

LONG		SHORT			
NOMINATIVE– ABSOLUTIVE	ERGATIVE	NOMINATIVE– ABSOLUTIVE	ERGATIVE	OBLIQUE- GENITIVE	
<i>aku</i>	<i>aku=</i>	—	<i>ku=</i>	<i>=ku</i>	‘1’
<i>kamu</i>	<i>kamu=</i>	—	—	<i>=mu</i>	‘2’
<i>engkau</i>	<i>engkau=</i>	<i>kau</i>	<i>kau=</i>	<i>(=kau)<sup>11</sup></i>	‘2’
<i>dia</i>	<i>dia=</i>	<i>ia</i>	<i>ia=</i>	<i>=nya</i>	‘3’

The personal pro-names *aku* and *engkau* have historical *d*-initial alternates *daku* and *dikau*, used after the prepositions *dengan* ‘with’ and *akan* ‘with regard to, about, for’, and after verb forms ending in the suffix *-kan* (one could formally consider it as a third case of the long form, the dative). But in the modern language, these *d*-alternates are practically no longer used. Historically, the third person pro-name long form *dia* may have been the *d*-alternate of *ia*.

The table shows that *aku* and *kamu* do not have a short nominative. A nominative *\*ku* or *\*mu* does not occur, as the following parallel examples demonstrate, where ‘X’ translates the variable pro-name:

- \*Ku bolèh baca surat ini.*
- \*Mu bolèh baca surat ini.*
- Kau bolèh baca surat ini.*
- Ia bolèh baca surat ini.*
- ‘X may read this letter.’

- \*Merèka minta agar ku membaca surat ini.*
- \*Merèka minta agar mu membaca surat ini.*
- Merèka minta agar kau membaca surat ini.*
- Merèka minta agar ia membaca surat ini.*
- ‘They ask that X read this letter.’

Similarly, *kamu* does not have a short ergative, and in Indonesian as it is spoken in Java and some other regions, *engkau* does not have a short oblique-genitive. In the following examples, which illustrate the implementation of the short form as ergative agent-referent, possessive attribute, direct and indirect object, respectively, the asterisk in parentheses before sentences with *=kau* reflects usage in some dialects of Indonesian, as mentioned:

<sup>11</sup> Attested in some dialects outside Java.

ergative agent-referent:	<i>Bolèh ku=baca surat ini?</i> * <i>Bolèh mu=baca surat ini?</i> <i>Bolèh kau=baca surat ini?</i> <i>Bolèh ia=baca surat ini?</i> 'May X read this letter?'
possessive attribute:	<i>Baik=lah sekarang dibaca surat=ku ini.</i> <i>Baik=lah sekarang dibaca surat=mu ini.</i> (* <i>Baik=lah sekarang dibaca surat=kau ini.</i> <i>Baik=lah sekarang dibaca surat=nya ini.</i> 'Well then let X's letter be read now.'
direct object:	<i>Merèka menyuruh=ku membaca surat itu.</i> <i>Merèka menyuruh=mu membaca surat itu.</i> (* <i>Merèka menyuruh=kau membaca surat itu.</i> <i>Merèka menyuruh=nya membaca surat itu.</i> 'They told X to read the letter.'
indirect object:	<i>Merèka minta kepada=ku agar membacakan surat ini.</i> <i>Merèka minta kepada=mu agar membacakan surat ini.</i> (* <i>Merèka minta kepada=kau agar membacakan surat ini.</i> <i>Merèka minta kepada=nya agar membacakan surat ini.</i> 'They ask X to read out this letter.'

For the third person, the table also shows how *ia* differs from *dia*. The latter may originally have been the *d*-alternate of the former, but there has apparently been a restructuring of the morphological paradigm. *Ia* cannot replace *dia* in the functions of direct or indirect object, or in that of possessive attribute, as the following examples show:

*Kata dia, ada yang akan memberi dia hadiah yang cocok untuk dia.*  
\**Kata ia, ada yang akan memberi dia hadiah yang cocok untuk dia.*  
\**Kata dia, ada yang akan memberi ia hadiah yang cocok untuk dia.*  
\**Kata dia, ada yang akan memberi dia hadiah yang cocok untuk ia.*  
'(S)he says, somebody will give her/him a present which is suitable for her/him.'

In all these positions, *ia* must be replaced by *=nya*:

*Kata=nya, ada yang akan memberi=nya hadiah yang cocok untuk=nya.*  
'(S)he says, somebody will give her/him a present which is suitable for her/him.'

The phrase translated as '(s)he says' literally means '[the]saying her/his', and thus has the third person pro-name as possessive attribute. It is noteworthy that *ia* is bisyllabic, whereas all other short forms in the personal nominals are monosyllabic (I will return to this below).

The oblique-genitive short form for the third person, *=nya*, is unique among all forms of the personal nominals, in that its use is not restricted to persons.

*Móbil saya perlu dibersihkan busi=nya.*  
'My car needs to have its sparkplugs cleaned.'

There also are grammaticalised cognates of *=nya*, functioning as definite article (in this it is grammatically closer to the demonstratives than to the personal articles), or as possessive copula; for instance,

*Busi=nya perlu diganti baru Pak.*

'The sparkplugs need to be renewed, Sir.'

*Atap=nya rumah mana yang bocor?*

'The roof of which house is leaky?'

Another particularity of *=nya*, one which it shares with *=mu*, is that it can also be used for a multiple referent.

*Rumah-rumah itu atap=nya bocor semua.*

'The roofs of all those houses are leaky.'

(lit. 'houses those roof its/their leak all')

*Hai kalian, ambil=lah barang-barang=mu masing-masing.*

'Hey, you, take each of your things.'

The defective or irregular distribution of the short forms of the personal pro-names is significant in that it helps certify that distinctions made here are not arbitrary impositions of eurocentric categories, but are indeed objectively immanent in the language on its own terms. The opposition of the oblique-genitive to the two other case forms is shown by the unique distribution of *=mu* and *=nya*. The opposition of the nominative-absolutive to the ergative is attested by the ergative *ku=* (as distinct from the oblique-genitive usage that *=ku* shares with *=mu* and *=nya*).

### 3 Relational pro-names

#### 3.1 Ambipersonal relational pro-names

The relational pro-names are words that derive from nouns that are either kinship terms or titles, but which functionally answer to the conventional notion of 'personal pronoun'. Grammatically, they have enough in common with the personal pro-names (including features that distinguish them from nonpersonal nominals) to place them with the personal nominals. Nevertheless, they also display features that distinguish them from the other personal pro-names, so that they must be considered as forming a subclass on their own.

The relational pro-names have been so labelled because, although they refer to some grammatical person, they do so in a relative rather than an absolute way. A relational pro-name indicates some social-status or kinship relationship between the speaker/writer and the listener/reader. Depending upon whether the implied social status fits the former or the latter of the interlocutors, it either expresses the first or the second person.

Hence, a sentence like *Kakèk sakit*. (lit. 'Grandfather is sick.') can mean either 'I am ill.' (elderly man speaking to youngster), 'You are ill.' (youngster speaking to elderly man), or 'Grandpa is ill.' (referring to neither of the interlocutors). In the last instance, the subject is not the relational pro-name but the noun from which it is etymologically derived. Although one could of course theoretically conceive it to be expressing the third person, it cannot appear as an ergative agent-referent because it is a noun, and hence a nonpersonal nominal.

Thus, in *Nasi itu sudah kakèk=makan* 'The rice has been eaten by X', 'X' can only be 'me' or 'you', never 'Grandpa' or even 'him'.

The relational pro-names usually also specify the gender of the referent, which is not the case with personal pro-names. Relational pro-names are divided into two subclasses: the ambipersonal relational pro-names, which can be used in alternation for first and second person, and the nonambipersonal ones, which only refer to the second person.

The ambipersonal relational pro-names derive from kinship terms. When not monosyllabic, they generally distinguish between long and short forms, listed below with a slash between them (when there is no short form, this is indicated with a dash). The chart gives the meaning of the original kinship term as the gloss, and then the social relationship of the referent person to the complementary speech partner.

AMBIPERSONAL RELATIONAL PRO-NAMES	KIN TERM MEANING	IMPLIED SOCIAL RELATIONSHIP
<i>kakèk/kèk</i>	'grandfather'	elderly, grey-haired male person, usually regardless of social status or rank, when complementary speech partner is not older than middle-aged
<i>embah/--</i>	'grandfather'	as above (particularly in Java)
<i>nènèk/nèk</i>	'grandmother'	female analogue of <i>kakèk</i>
<i>bapak/pak</i>	'father'	male person of next older generation or of correspondingly higher social status; senior male public figure
<i>ayah/--</i>	'father'	father
<i>babèh/bèh</i>	'father'	father (in Jakarta)
<i>papa/pap</i>	'father'	father (mainly in urban environment)
<i>papi/pap</i>	'father'	as above
<i>ibu/bu</i>	'mother'	female analogue of <i>bapak</i>
<i>mama/mam</i>	'mother'	female analogue of <i>papa</i>
<i>mami/mam</i>	'mother'	female analogue of <i>papi</i>
<i>embok/bok</i>	'mother' <sup>12</sup>	adult female servant, labourer, peasant, artisan, flying saleswoman, all regardless of age difference
<i>uak/wak</i>	'senior uncle'	male person of next older generation, older than one's parents (mainly in familiar environment)
<i>pakdé/--</i>	'senior uncle'	(in Java equivalent to <i>uak</i> )

<sup>12</sup> Only used in this meaning in some regions in Java. In environments where it occurs as a neutral-style kinship term, the pro-name can be used in the same way as *ibu*. Otherwise, the use with reference to persons not pertaining to the indicated particular social group may have a pronounced pejorative connotation. Some relatively undemocratic attitudes in traditional usage are described here as objectively observed, without any implications of a commendatory or evaluative sense from the author. It should furthermore be borne in mind that usage is regionally, dialectally and stylistically variable. It is also very much in flux.

<i>mamak/mak</i>	'junior uncle'	in Sumatra male person of next older generation, younger than one's parents (mainly in familiar environment)
<i>paklék/--</i>	'junior uncle'	(in Java equivalent to <i>mamak</i> )
<i>paman/--</i>	'uncle'	male person of next older generation (mainly in familiar environment); translates foreign-language 'uncle' before proper name
<i>oom</i> <sup>13</sup>	'uncle'	male person of next older generation (mainly in familiar urban environment)
<i>budé/--</i>	'senior aunt'	female analogue of <i>pakdé</i>
<i>bulék/--</i>	'junior aunt'	female analogue of <i>paklék</i>
<i>bibi/bi</i>	'aunt'	female analogue of <i>paman</i>
<i>tante/--</i>	'aunt'	female analogue of <i>oom</i>
<i>kakak/kak</i>	'elder sibling' <sup>14</sup>	person of the same generation but older or senior in social rank regardless of gender; boyfriend or husband, regardless of age difference
<i>abang/bang</i>	'elder brother'	male person of the same generation but older in age or senior in social rank; boyfriend or husband, <sup>15</sup> adult male popular public figure with intellectual stature; adult male servant, labourer, peasant, artisan, flying salesman, regardless of age difference <sup>16</sup>
<i>bung</i>	'elder brother' <sup>17</sup>	male person of the same generation but older in age or senior in social rank; adult male popular public figure with political stature
<i>mas</i>	'elder brother'	male person of the same generation but older in age or senior in social rank (particularly in Java)
<i>kakang/kang</i>	'elder brother'	(in West Java equivalent to <i>mas</i> , in familiar environment)
<i>cacak/cak</i>	'elder brother'	(in East Java and Madura equivalent to <i>mas</i> , in familiar environment)
<i>embakyu ~ embak</i> <sup>18</sup>	'elder sister'	female analogue of <i>mas</i> ; adult female popular public figure

<sup>13</sup> Pronounced as if spelled *óm*.

<sup>14</sup> In some regions, the kinship term *kakak* means 'elder sister' and the relational pro-name refers only to female persons.

<sup>15</sup> Mainly in regions where *kakak* means 'elder sister'.

<sup>16</sup> As pro-name for the last mentioned group of referents, *abang* appears to serve as the male match to *embok*. Nevertheless, persons of this social group can also be referred to as *bapak* or *ibu*, depending upon the gender.

<sup>17</sup> Only apparently used in this meaning in some regions in Maluku.

<sup>18</sup> When the shorter form of a relational pro-name is not monosyllabic, it retains the grammatical properties of the long form. In other words, it is not a short form in the grammatical sense. In the speech of persons having Javanese as their first language, *embakyu* is usually pronounced as in Javanese, that is, with an initial

<i>uni/--</i>	'elder sister'	(in West Sumatra older female person of same generation)
<i>adik/dik</i> <sup>19</sup>	'younger sibling'	person of same generation but younger in age or lower in social rank regardless of gender; girlfriend/wife
<i>ajeng/jeng</i>	'younger sister'	(in Central Java younger female person of same generation)
<i>anak/nak</i>	'child'	person of next younger generation
<i>cucu/cu</i>	'grandchild'	child or youthful person when complementary speech partner is in the social class of 'grandparent'

Region-specific terms have only been included when their use is frequently encountered outside the region or in ethnically mixed environments. Most of the included also occur in literary works of national culture relevance.

The traditional use of many of these pro-names is such that they form complementary pairs. A person would then use one term of the pair for '1', and the other one for '2'. His or her dialogue partner would then use the same pair but with opposite meanings. A well-known dialogue from James Fenimore Cooper would then go like this:

Cora: *What would <elder brother> say to <younger sibling>, the daughter of Munro?*

Magua: *<Elder brother> was born a chief and a warrior among the red Hurons of the lakes; <elder brother> saw the suns of twenty summers ... before he ever saw a pale-face.*

Here <Elder brother>, whether *kakak*, *abang* or *mas*, would in turn translate as 'you' (Cora speaking), and then as 'I' (Magua speaking), and the same would apply to <younger sibling> *adik*.

Such complementary pairs traditionally consisted of semantically reciprocal terms, but this is no longer very strictly adhered to. Thus pairs like *uak* to *anak* ('uncle' to 'child'), or even *ibu* to *adik* ('mother' to 'younger sibling') can sometimes be observed. The conversing persons do not necessarily have to use the matching pairs. But if they don't use them, one of the interlocutors will usually not be using relational pro-names, but some other personal nominals (either personal pro-names or proper names).

In addition to the traditional use in pairs of reciprocal terms (semantically logical or otherwise), one can also observe speakers addressing each other by the same term, both implying the other's seniority, as when both speakers refer to each other as *mas* or *bapak*. In this situation, in which the relational pro-name is only used for '2', a personal pro-name will often be used for '1' (e.g. *saya*—in formal situations *kami*).

---

prenasalised bilabial stop, making it bisyllabic. In Indonesian *embak* is otherwise trisyllabic, having either *e* or a syllabic bilabial nasal as the nucleus of the first syllable. Dialectal difference in syllabicity of the shorter form *embak* seems to have led to a third, even shorter form, which will be transcribed here as *bak*, in the speech of speakers with 'non-Javanese' pronunciation. This latter, and the 'Javanese' monosyllabic realisation of *embak* (often spelled *mbak*) represent the grammatical short form.

<sup>19</sup> Pronounced as if spelled *adék*, with final *k* realised as a glottal stop.

Some of the ambipersonal relational pro-names have an honorific derivation formed with the suffix *-(a)nda* (with sandhi when the basic form ends in *k*, resulting in deletion of the latter), expressing either respect or endearment. These are *ayahanda*, *ibunda* or *bunda*, *ananda* or *nanda*, *kakanda* or *kanda*, *adinda* or *dinda*. The two last-mentioned pairs are mainly used for boyfriend/husband and girlfriend/wife, respectively. The shorter forms, not being monosyllabic, figure grammatically as variants of the long form, rather than as short forms.

The ambipersonal relational pro-names have three cases—nominative-absolutive, ergative and vocative—and an article mode. Long and short forms only occur in the vocative case and the article mode, whereas the nominative-absolutive and the ergative of the ambipersonal relational pro-names are always long. The article mode has been so labelled because it functions like the personal articles *si=* and *sang*. In the relational pro-names, the short form in the article mode is always a proclitic. Some of the ambipersonal relational pro-names (e.g. *anak*, *ayah*, *papa*, *mama*) do not appear to have an article mode, but usage may vary dialectally in this respect.

The following table displays the paradigm of some frequently used relational pro-names, illustrating at the same time the existing variety in distribution patterns.

**Table 3:** Frequently used relational pro-names

NOMINATIVE–ABSOLUTIVE	ERGATIVE	VOCATIVE		ARTICLE MODE	
		LONG	SHORT	LONG	SHORT
<i>ibu</i>	<i>ibu=</i>	<i>ibu</i>	<i>bu</i>	<i>Ibu</i>	<i>Bu=</i>
<i>bapak</i>	<i>bapak=</i>	<i>bapak</i>	<i>pak</i>	<i>Bapak</i>	<i>Pak=</i>
<i>anak</i>	<i>anak=</i>	<i>anak</i>	<i>nak</i>	—	
<i>embakyu</i> ~ <i>embak</i>	<i>embakyu=</i> ~ <i>embak=</i>	<i>embakyu</i> ~ <i>embak</i>	<i>bak</i>	<i>Embakyu</i> ~ <i>Embak</i>	<i>Bak=</i>
<i>mas</i>	<i>mas=</i>		<i>mas</i>		<i>Mas=</i>
<i>adik</i>	<i>adik=</i>	<i>adik</i>	<i>dik</i>		<i>Dik=</i>

The distinctive use of the nominative-absolutive and the ergative is parallel to that of personal pro-names; compare (mother or older lady speaking):

Nominative-absolutive:

*Ibu belum pernah jual.*

‘I have never sold.’

*Ibu belum pernah jual buku.*

‘I have never sold books.’

*Buku, ibu belum pernah jual.*

‘Books I have never sold.’

*Buku itu dijual oleh ibu.*

‘That book was sold by me.’

*Buku ibu belum dijual.*

‘My book has not yet been sold.’

Ergative:

*Belum pernah ibu=jual, buku itu.*

'I have never sold that book.'

*Buku itu belum pernah ibu=jual.*

'That book has never been sold by me.'

The grammatical distinctness of the vocative is attested in that this is the only case in which there is a short form. Thus in the above six glosses it is not possible to replace *ibu* (or *ibu=*) by *bu* (or *bu=*). However, for some speakers the ergative also apparently has a short form, and replacement with *bu=* would be possible in their speech in the two examples above. Compare also the following, with someone speaking to a younger person of the same generation or to a social junior. (*Dik* before *Katrin* demonstrates the article mode, placing Catherine in the same social category.)

*Adik, apa=kah adik sudah kenal dengan Dik=Katrin?*

*Dik, apa=kah adik sudah kenal dengan Dik=Katrin?*

*\*Adik, apa=kah dik sudah kenal dengan Dik=Katrin?*

*\*Dik, apa=kah dik sudah kenal dengan Dik=Katrin?*

'Say, are you already acquainted with [Sis] Catherine?'

And, for the ergative case:

*Tolong adik=bacakan surat itu.*

*(\*)Tolong dik=bacakan surat itu.*

'Please read [me] the letter.'

(That the 'reading' in the above ergative case example is done for the benefit of someone—in this case for 'me'—is signalled by the verbal suffix *-kan* and does not bear on the matter under discussion; *adik* means 'you'.)

When the monosyllabic form is the only one, there is no such constraint, thus demonstrating that the feature is not (or is no longer) phonologically or prosodically conditioned.

*Mas, apa=kah mas sudah kenal dengan Mas=Udin?*

'Say [brother], are you already acquainted with Udin?'

*Tolong mas=bacakan surat itu.*

'Please read [me] the letter.'

The relational pro-name in the vocative can be part of a phrase or group expressing an appeal or invitation, and then it typically does not stand at the beginning:

*Mari bu, tamu lain sudah menunggu ibu.*

'Come on, ma'am, the other guests are already waiting for you.'

*Awas batu=nya bu, jangan sampai ibu tersandung.*

'Beware of the stones, ma'am, lest you may trip.'

The use of the article mode will be discussed in greater detail in the section on the personal articles. A peculiarity of the article mode of the ambipersonal relational pro-names is that it generally has both a long and a short form. The short form is actually the more frequently used one.

*Pagi ini, Bak=Susi dan Mas=Petrus pergi menjenguk Bu=Ijah dan Pak=Ali.*

'This morning, Suzy and Peter went to visit [Ma] Ijah and [Pa] Ali.'



One reason for the preference for the short form appears to be that it helps disambiguate two common kinship-term constructions: kin-term plus proper-name equivalent (where the kin-term is a pro-name) vs kin-term plus proper-name possessor (where it is the noun). (As a nonpersonal nominal, the noun does not normally have a short form, but even when it has, the short form is not grammatically distinct from the longer variant.) Compare the following, in which the nature of a possessive attribute can always be tested by inserting the optional possessive copula =*nya*.

*Ini Adik Katrin.*

*Ini Dik=Katrin.*

'This is [younger sister] Catherine.'

*Ini adik[=nya] Katrin.*

*\*Ini dik[=nya] Katrin.*

'This is Catherine's younger sister.'

The long form in the article mode, as in the second of the two following examples, usually expresses greater respect.

*Pak, bagaimana sikap Pak=Menteri terhadap usul baru itu?*

*Pak, bagaimana sikap Bapak Menteri terhadap usul baru itu?*

'Sir, what is your [the Minister's] attitude towards the new proposal?'

The nonreduplicated forms of the ambipersonal relational pro-names all refer to a single person, unlike the respective kinship-term nouns, in which the unreduplicated form denotes an unspecified number of referents when not specified in the context. The unreduplicated form of the pro-name can therefore be called the singular. The relational pro-names also have a plural, which is formed by reduplication, for example:

*Anak-anak, hati-hati=lah jangan sampai anak-anak berenang ke tempat yang dalam.*

'Children, be careful that you don't swim to the deep place.'

*Bapak-bapak dan ibu-ibu, silahkan hidangan di atas meja itu bapak-bapak=dan=ibu-ibu=icipi semua=nya.*

'Ladies and gentlemen, please take a taste of all the dishes that are laid out on the table.'

The latter sentence shows a somewhat intricate example of an ergative construction, in which the entire group connected with equal signs is pronounced like a single word.

### 3.2 Nonambipersonal relational pro-names

The nonambipersonal relational pro-names do not differ very significantly from the ambipersonal, except that they are not used for reference to the first person. On the semantic plane, they usually specify gender and relative social status, as do most of the ambipersonal pro-names. With regard to expression of social relationships, usage often still reflects long-standing relics of feudal social stratification. On the morphological plane the nonambipersonal relational pro-names likewise distinguish three case forms (absolute, ergative, vocative) and an article mode.

Some frequently used nonambipersonal relational pro-names follow. The list could be further lengthened, because several more nouns denoting persons can be converted into pro-names and thus become terms of address with reference to an individual person.

NONAMBIPERSONAL PRO-NAMES	LITERAL MEANING	IMPLIED SOCIAL RELATIONSHIP
<i>saudara/--</i>	'sibling, relative'	formal reference to indigenous male adult; lately also to foreigners acquainted with indigenous custom
<i>saudari/--</i>	'female relative' <sup>20</sup>	female analogue of <i>saudara</i>
<i>tuan/--</i>	'sir, Mr'	used before 1945–50 in formal reference to middle- or upper-class male adult, then only to foreign male adult, or by blue-collar personnel to male employer or customer. Since the 1970s, again used more generally to indigenous male adult customer or business client
<i>juragan/gan</i>	'sir, Mr'	(in West Java equivalent to <i>tuan</i> )
<i>ndoro/--</i>	'sir, Mr'	(in East and Central Java equivalent to <i>tuan</i> )
<i>nyonya/nyah</i>	'madam, Mrs'	female analogue of <i>tuan</i>
<i>nona/non</i>	'miss'	formal reference to young female foreigner or customer/client, or to daughter of foreigner/customer/client/employer; ironic but usually friendly reference to a girl
<i>nènèng/nèng</i>	'miss'	(in West Java equivalent to <i>nona</i> )
<i>sinyó/nyó</i>	'young master'	reference by blue collar personnel to son of customer/employer before 1945, since then mainly to son of foreigner; ironic pejorative reference to boy or young man as 'preppy boy'
<i>upik/pik</i>	'little girl'	(in West Sumatra female child)
<i>buyung/--</i>	'little boy'	male analogue of <i>upik</i>
<i>ujang/jang</i>	'little boy'	(in West Java) as above
<i>sus</i> <sup>21</sup>	'sister'	female person of the same generation (mainly in urban environment)

The nonambipersonal relational pro-names that do not derive from kinship terms do not seem to have a short form in the article mode. Whether derived from kinship terms or not, some of the nonambipersonal pro-names do not seem to be used in the article mode.

<sup>20</sup> Artificially formed as a feminine counterpart to *saudara* (which originally referred to siblings of either gender) by analogy to the pair of Sanskrit loans *putra* 'son, prince' and *putri* 'daughter, princess'.

<sup>21</sup> Pronounced as if spelled *ses*. Etymologically, this is short for *suster* (from Dutch *zuster* 'sister'; the Indonesian is usually pronounced as if spelled *sester*). However, *suster* is only used with reference to missionary and medical sisters. The corresponding relational pro-name is also nonambipersonal, but it does not function as the long correspondent of a short *sus*. Persons who could be referred to by the one (either *sus* or *suster*), could as a rule not be alternatively referred to by the respective other. The formation of the short form took place in Dutch (*zuster* → *zus* 'sis'), and each of the two forms appears to have been borrowed into Indonesian Malay independently. It is also common to meet both the long and the short word spelled as in Dutch (with initial z-) in usage by Indonesian intellectuals, but this is seldom reflected in the pronunciation.

**Table 4:** Nonambipersonal relational pro-names

NOMINATIVE- ABSOLUTIVE	ERGATIVE	VOCATIVE		ARTICLE MODE	
		LONG	SHORT	LONG	SHORT
<i>saudara</i>	<i>saudara=</i>	<i>saudara</i>	--	<i>Saudara</i>	
<i>tuan</i>	<i>tuan=</i>	<i>tuan</i>	--	<i>Tuan</i>	
<i>nyonya</i>	<i>nyonya=</i>	<i>nyonya</i>	<i>nyah</i>	<i>Nyonya</i>	
<i>ndoro</i>	<i>ndoro=</i>	<i>ndoro</i>	--	--	
<i>nènèng</i>	<i>nènèng=</i>	<i>nènèng</i>	<i>nèng</i>	<i>Nènèng</i>	<i>Nèng=</i>
<i>upik</i>	<i>upik=</i>	<i>upik</i>	<i>pik</i>	--	--
<i>sus</i>	<i>sus=</i>		<i>sus</i>		<i>Sus=</i>

The syntactic constraints on the use of the short form of those ambipersonal relational pro-names that have long and short forms also apply to the nonambipersonal ones. In the standard language it is restricted to the vocative case and the article mode, but for some speakers it is additionally represented in the ergative.

*Nyonya, apa=kah bèsok nyonya mau memakai móbil?*

*Nyah, apa=kah bèsok nyonya mau memakai móbil?*

\**Nyonya, apa=kah bèsok nyah mau memakai móbil?*

\**Nyah, apa=kah bèsok nyah mau memakai móbil?*

'Ma'am, will you be wanting to use the car tomorrow?'

*Móbil itu apakah bèsok mau nyonya=pakai?*

(\**Móbil itu apakah bèsok mau nyah=pakai?*

'Will you [madam] be wanting to use the car tomorrow?'

Here too, the constraint is lifted when the monosyllabic form is the only form.

*Sus, kalau sus mau, móbil=nya bisa sus=pakai se=harian.*

'Miss, if you want, you can use the car the whole day.'

The unreduplicated form of nonambipersonal relational pro-name refers to a single person. It forms the plural by reduplication in the same way as the ambipersonal:

*Saudara-saudara, mari=lah saudara-saudara=periksa ketepatan letak barang-barang itu.*

'Gentlemen, do come and check the correctness of the placing of those objects.'

The pro-names *saudara* and *saudari* are peculiar in that they form a mixed plural in a manner similar to the formation of collective nouns<sup>22</sup>. (This usage has existed since long before 'political correctness' became a public issue internationally.)

*Saudara-saudari, mari=lah saudara-saudari=periksa ketepatan letak barang-barang itu.*

'Ladies and Gentlemen, do come and check the correct placing of those objects.'

<sup>22</sup> Collective nouns composed from pairs of complementary kinship terms, such as *bapak-ibu* 'parents (lit. father-mother)', *kakak-adik* 'siblings (lit. elder sibling-younger sibling)' are not generally converted into semantically corresponding personal pro-names.

#### 4 Personal proper names

Personal proper names in Indonesian differ from other proper names (particularly topological<sup>23</sup> and astronomical) in that they share grammatical properties of the pro-names. Like a pro-name, a personal proper name refers to a single individual. But unlike the former, it does not seem to have a reduplicated plural. However, a personal proper name can be converted into a noun, upon which it acquires a plural, while its nonplural form is nonspecific in number. As a noun, it also can combine with a numerical classifier (in this case the classifier for persons *orang*). This special usage is illustrated in the following sentences, in which *Ali* is not a personal proper name in the grammatical sense. It carries a different grammatical meaning than its personal nominal cognate.

*Di kelas kami ada tiga [orang] Ali.*  
'There are three Ali's in our class.'

*Semua Ali-Ali harap berdiri!*  
'Let all the Ali's stand up!'

Personal proper names in Indonesian can be used in the same way as ambipersonal relational pro-names. This usage at the same time implies inclusion of the speech participants within a familiar circle, and thus establishes a particular social relationship between speaker and listener. In the following example (a person named Yono speaking to person named Ali), *Ali* appears in its original role as a personal proper name, once in the vocative and once in the ergative. (Nouns allow neither possibility.)

*Ali, tolong Ali=perkenalkan Yono dengan Yanti.*  
'Ali, please introduce me to Yanti.'

Abridgment is more elaborate in the personal proper names than in the other personal nominals. Long personal proper names may have several polysyllabic shorter variants that nevertheless figure as long forms in the grammatical sense. Only the (usually, but not always uniquely) monosyllabic shorter variant functions as a short form in the grammatical sense. Among the long forms one can thus distinguish a full form and one or more polysyllabic reduced forms on an etymological, not a grammatical, basis. Moreover, most proper names also have a "restored" or "reconstructed" long form, achieved by full or partial reduplication of the monosyllabic short form. Table 5 shows some typical examples, illustrating also the treatment of foreign names. (A slash is used between alternative variants, and a wave sign between variant spellings of an identical mode.<sup>24</sup>)

<sup>23</sup> This includes geographical names such as *Indonesia*, *Sumatra*, *Jakarta*, *Ciliwung* (a river), *Krakatau* (a volcano), as well as names of streets, architectural or monumental objects, landmarks etc., such as *Malióboró* (street in Yogyakarta), *Pejambon* (building of the foreign ministry), *Ki Amuk* (old canon in Banten).

<sup>24</sup> Forms which are spelled with final *y* and *ie* (the latter also before final *k*) are read as *i*. Forms which are spelled with a final vowel are often pronounced with a following glottal stop. This is sometimes indicated in a variant spelling with final *k*. A missing final *k* in the spelling does not generally exclude the variant pronunciation.

**Table 5:** Typical forms of proper names

LONG			SHORT
FULL	REDUCED	RESTORED	
<i>Artati</i>	<i>Tati ~ Tatie ~ Tatiek</i>	<i>Titi ~ Titiek</i>	<i>Ti ~ Tiek</i>
<i>Khotijah</i>	<i>Ijah</i>	<i>Iid</i>	<i>Id / Jah</i>
<i>Irawan</i>	<i>Iwan</i>	<i>Wawan</i>	<i>Wan</i>
<i>Syarifuddin</i>	<i>Udin</i>	<i>Didi</i>	<i>Di / Din</i>
<i>Mohammad</i>	<i>Amat</i>	<i>Mamat</i>	<i>Mat</i>
<i>Bènyamin</i>	<i>Bènnny ~ Bènnie ~ Bèni</i>	—	<i>Bèn</i>
<i>Joni</i>	—	—	<i>Jon</i>
<i>Bèng An</i>	—	<i>Aan</i>	<i>An</i>

Note that *Iid* and *Aan* (and many other similar proper-name forms reconstructed by reduplication of the short form) have an intervocalic glottal stop that is not noted in the spelling. In the case of personal proper names like *Bambang* (*Bang* for short), a restored bisyllabic form would be identical with the etymological full form.

It is also noteworthy that a monosyllabic foreign name like English *John* would be treated as the short form of a perhaps extant polysyllabic diminutive, in this case *Johnny*. This applies, for instance, to the Indonesian name *Joni*, which is in fact a borrowing from English. An analogous case is *Yópi* and its short form *Yóp* (from Dutch *Joopie/Joopje*, diminutive of *Joop*, in Dutch the short form of *Joseph*). The relationship between long and short forms of borrowed personal proper names in Indonesian is thus determined by syllabicity alone, and is quite independent of the relationship of the respective etyma in the donor language.

It frequently happens in familiar circumstances that a person, particularly a small child, is called by some monosyllabic nickname that is not derived from his or her ‘correct’ proper name. In that case, only the restored form is sometimes available as a grammatical long form. In urban environments, particularly with Indo-Eurasian contacts, the formant of the diminutive in Dutch (the suffix *-tje*), borrowed into Indonesian as *-ce*, is sometimes implemented to reconstruct a (bisyllabic) long form from the short one.

long *Dini* → short *Din* → long *Dince*

Table 6 shows the distribution of grammatical forms of two personal proper names. It is already familiar from that of the relational pro-names, save for the circumstance that the proper names do not have an article mode.

**Table 6:** Grammatical forms of personal proper names

NOMINATIVE-	ERGATIVE	VOCATIVE	
		LONG	SHORT
<i>Khotijah / Ijah / Iid</i>	<i>Khotijah= / Ijah= / Iid=</i>	<i>Khotijah / Ijah / Iid</i>	<i>Id / Jah</i>
<i>Irawan / Iwan / Wawan</i>	<i>Irawan= / Iwan= / Wawan=</i>	<i>Irawan / Iwan / Wawan</i>	<i>Wan</i>

Compare the following examples, in which a short vocative form in the address can be replaced by long a form except where asterisked.

*Wan, kamar Irawan apa=kah sudah Wawan=benah?*

*Wan, kamar Iwan apa=kah sudah Iwan=benah?*

*\*Wan, kamar Wan apa=kah sudah Wawan=benah?*

*\*Wan, kamar Iwan apa=kah sudah Wan=benah?*

'Irawan, have you tidied up your room?'

*Jah, apa=kah Ijah nanti bisa tolong Wawan membenah kamar?*

*\*Jah, apa=kah Jah nanti bisa tolong Wawan membenah kamar?*

*\*Jah, apa=kah Ijah nanti bisa tolong Wan membenah kamar?*

'Ijah, could you later help me tidy my room?'

## 5 Personal articles

The personal articles could be considered as pro-names having a defective paradigm of forms, consisting only of the article mode. There are three in the modern language:<sup>25</sup>

*si=* ~ common singular personal article (sometimes pejorative);

*sang* ~ honorific singular personal article;

*para* ~ plural personal article (intermediate in degree of honorification, compared with the two former)

A personal article personalises the referent of the word that follows it, impressing upon it at the same time the quality of individual distinctness (in the case of the singular articles: also that of singularity) and a certain social status. However, in modern usage, this latter effect of the articles has become much eroded. Thus, socially neutral use of *si=* and *para* has become relatively widespread. Correspondingly, the use of *sang* has lost much of its productivity, being mainly (but not exclusively) limited to some fixed expressions such as:

*sang raja* 'the king' (mainly with reference to an historical monarch)<sup>26</sup>

*Sang Dwiwarna* 'The Bicolor' (the Indonesian red-and-white national flag)

*sang isteri* 'the Missus' (mildly ironic reference to one's own or another's wife)

<sup>25</sup> In historical texts, there additionally is *hang* for men's names, and *dang* for women's, both being also honorific. They now only appear in names of certain historical or legendary figures, and one could consider them as fixed expressions.

<sup>26</sup> Does not translate as 'His Majesty the King', for which the corresponding Indonesian is *Baginda Raja*.

The common singular correspondent *si=* has several uses. It is traditionally placed before a personal proper name referring to a person of comparable or lower social standing than the speaker.

*Si=Didi dan Si=Aminah sedang pergi ke sekolah.*  
'Didi and Aminah are going to school.'

This usage is on the decrease compared with half a century ago and earlier. It is fully analogous to the use of the article mode of the personal pro-names.

*Dik=Didi dan Dik=Aminah sedang pergi ke sekolah.*  
'Didi and Aminah are going to school.'

The gradual decrease of such usage seems to be conditioned by a trend towards more informality, in which explicit expression of the social relationship decreases in importance, and by a trend towards economy achieved by simply dropping the article:

*Didi dan Aminah sedang pergi ke sekolah.*  
'Didi and Aminah are going to school.'

Both the personal article and the article mode of pro-names can be used to personalise animals and other nonhuman figures in fables and fairy tales, as in *Si=Kancil* 'Mousedeer', *Bung=Kelinci* 'Br'er Rabbit'. Before the name of a person of higher social rank, the use of *si=* is explicitly pejorative. It is also perceived as inappropriate before the name of a person of equal rank who would expect formal address. As a consequence, *si=* can sometimes be used to express familiarity.

Whereas the use with proper names is on the decrease, *si=* is increasingly being used as a 'definite article' for terms denoting persons other than proper names. Interesting is the use before some pro-names:

*si=dia* 'her/his one-and-only' (lit. 'the (s)he', ironic reference to someone's sweetheart)

*Si=Bung* 'the Bung' (ironic reference to person normally addressed *bung*; when that is a person of higher standing, such reference is only possible under closely familiar circumstances)

Analogous to the latter are *si=abang*, *si=buyung*, *si=embak*, *si=nènèng*, *si=nona*, *si=ujang*, etc. The irony expressed here is somewhat similar to that in *sang isteri* 'the Missus', but involves a greater degree of familiarity, and does not include the 'grudging respect' acknowledged in *sang isteri*. This special use of *si=* is not attested for the article mode of relational pro-names.

Another use of *si=* which is on the increase is that before deverbal nominals, derived by prefixation of *peN-*. This stresses the nominality of the derived word, and (as a rule) indicates that the referent is a person (that is, not an inanimate instrument, for example). It also has the effect of establishing singularity of the referent (not implied by the noun form itself).<sup>27</sup>

<sup>27</sup> When plurality needs to be expressed, *si=* is replaced by *para*. In both instances, number is always explicit, i.e. never unspecified as in the use of the noun in the absence of an article.

*si=pengirim* 'the sender' (e.g. of a letter;  *kirim* 'send')

*si=penulis* 'the writer' (*tulis* 'write')

*si=pembaca* 'the reader' ( *baca* 'read')

In this, no familiarity or pejority is usually expressed. Exceptions mainly arise when the derived word is a term which has tended to become a direct reference to a person. Thus, in a generalised context, it is alright to say *si=pemimpin* 'the leader' (*pimpin* 'lead'). But in a discussion about a particular high-ranking person who may sometimes be referred to by the word 'leader', the expression could acquire a pejorative undertone. Nevertheless, the sender of a letter, an author or a reader may be referred to by one of the previous three examples even if he or she is a very high-ranking, revered or august personality. The use of *sang* in place of *si=* when the referent is high-ranking would render the reference lightly ironic. But this is probably not so much in consequence of the inadequacy of the honorification, as of the relative obsolescence of *sang*.

Special are *si=pelaku* 'the actor, performer' and *si=penderita* 'the sufferer' when used as grammatical terms meaning 'agent' and 'patient' respectively, in which case they do not necessarily refer to persons or to singular referents.

The plural personal article *para* originally implied adherence of the referent persons to the upper class, as in *para pembesar* 'the dignitaries' or *para tamu* 'the guests'. But usage has gradually liberalised on this point, and it is no longer impossible to write *para petani* 'the peasants' or *para penyamun* 'the robbers', although the latter may have a lightly ironic connotation.

On the semantic plane, the difference between expression of plurality with the plural personal article (e.g. *para pemimpin* 'the leaders'), and with the normal form of the plural (*pemimpin-pemimpin [itu]* '[the] leaders')—particularly when definiteness in the latter instance is expressed with additional means—seems to lie mainly in that the normal plural is more matter-of-fact, whereas the complex with personal articles implies a particular interest in or individuality of each of the persons collectively referred to.

## 6 Pro-name groups

The term 'pro-name group' is used here to denote a nominal group occupying as a whole the position of a pro-name in a sentence. There seem to be two basic mechanisms by which such a group can be formed.

The first is by reference to oneself (first person) or the conversation partner (second person) by a group of words. This is comparable with the conversion of a noun into a relational pro-name. Actually, this latter procedure could be regarded as the formation of a pro-name group consisting of only one word. The group as a whole can be in the vocative, absolutive or ergative case (the problem of case forms of word groups will be addressed separately below). In the following examples, the speaker is addressing *tamu-tamu=ku yang terhormat* 'my respected guests':

Vocative:

*Tamu-tamu=ku yang terhormat, mari kita pindah ke kamar makan.*

'My respected guests, let us move to the dining room.'



Nominative-absolutive:

*Silahkan=lah tamu-tamu=ku yang terhormat memilih sendiri tempat duduk masing-masing.*

'Please each of you [my respected guests] choose your own seat.'

Ergative:

*Katakan=lah, gambar mana=kah yang paling tamu-tamu=ku=yang=terhormat=senangi.*

'Do tell me, which of the pictures you [my respected guests] like the most.'

*Anggur yang baru tamu-tamu=ku=yang=terhormat=minum ini dibuat di Australia.*

'The wine that you [my respected guests] have just drunk is produced in Australia.'

Regardless of length, the whole pro-name group in the ergative is pronounced as one word together with the verb form following it. This is perhaps comparable, at least with regard to the prosodic phonology, with compounds like *the coming-and-going* in English.

The other way of forming a pro-name group is by simply preposing a personal article or the article mode of a relational pro-name before another word which does not need to be a personal nominal. Some examples were already given in the section on personal articles. The following will merely illustrate the procedure in the ergative case:

*Orang-orangan yang telah Bung=Kelinci=tinju itu terbuat dari tèt.*

'The dummy which you [Br'er Rabbit] have socked is made of tar.'

*Dapat=kah kira=nya pelanggaran kecil ini Pak=Pólisi=maafkan untuk se=kali ini?*

'Could you [Mister Police Officer] perhaps excuse this small transgression for this one time?'

One final example illustrates a combination of the two, in which the speaker refers to himself as *pak=guru=mu ini* 'your teacher here':

*Anak-anak, kalau anak-anak rajin, anak-anak nanti pak=guru=mu=ini=berikan angka baik.*

'Children, if you are diligent, I'll give you good marks.'

## 7 Conclusions

A closer examination of the lexical subclasses grouped together into the class of personal nominals (leaving aside the personal articles) shows, in my opinion, that they indeed share a number of characteristic features that in turn collectively contrast them with subclasses grouped together into the class of nonpersonal nominals. The personal articles, though not being nominals in a strict sense, have only been included here because they correspond quite closely to the article mode of the pro-names. They could thus be seen as defective pro-names.

On the semantic plane, the personal nominals denote single individual referents, except for some personal pro-names which refer to groups of persons. Nouns, on the other hand, are nonspecific in number when not marked as plural, and thus do not explicitly denote single referents.

Another feature common to the personal nominals is that they express a certain social relationship between the speech participants. This feature seems to reflect social traditions still prevalent in the speech community, which, although gradually on the way out, have far

from disappeared. One may speculate that the language feature will probably disappear together with the social feature.<sup>28</sup>

Noteworthy are conscious efforts to overcome reflections of social stratification in the language by abandoning the use of such relational pro-names as *tuan* 'Mr' and *nyonya* 'Mrs' in favour of *saudara* and *saudari* respectively, inspired by the historical use of *citoyen(ne)* during the French Revolution. The failure of such quite sympathetic and well-meant reform efforts in language usage to remedy the social situation may perhaps be seen as a reminder that language merely serves as a means of expressing relationships existing either in reality or imagination, and abolishing a reference naturally does not abolish the referent.

On the grammatical plane, personal nominals are contrasted with nonpersonal nominals in that the former can function as agent reference in ergative constructions. In all the classes of personal nominals other than the personal articles, there thus is a distinction between an ergative case form and the nominative-absolutive (which is in turn split into nominative-absolutive and oblique-genitive in short forms of a few personal pro-names). The relational pro-names and the personal proper names furthermore distinguish a vocative case.

Another apparently characteristic feature shared by most, but not all, personal nominals is the opposition of long and short forms, which has been shown to have syntactic relevance. This too is a feature not found in the nonpersonal nominals. Particularly interesting in this regard is the 'reconstruction' of a variant long form from the short form in the personal proper names. The constraints in the use of the short form serve as an important indicator that assumptions made above—that certain distinctions were grammatically relevant—do indeed reflect the objective structure, and are not description artefacts.

Of a more fundamental nature is perhaps the question of just how far the description of various syntactic placements of the personal nominals as case forms is justified. With partial exception of the short forms of the personal pro-names, the respective forms are practically homonyms. The only phonologically palpable marking is clisis, the ergative forms being proclitics. But this too seems more likely to be a syntactic rather than morphological feature.

The suspicion may easily arise, therefore, that 'eurocentric' concepts of noun case are being imposed upon morphologically amorphous nominals. Yet phonological marking is not a required criterion for distinction of forms within the morphological paradigm, and a well-known counter-example is that of the present participle and the gerund in English, both being formed by adding the suffix *-ing* to the base and being homonyms for all English verbs, regular as well as irregular.

A theoretical foundation for drawing such distinctions in the description of Indonesian was proposed in Mahdi (1993:182). It was indicated, that contrast in characteristic syntactic valencies had up to then served as the basis for distinguishing word classes in isolating languages or in languages with a high frequency of conversion (phonologically unmarked word derivation). It was noted, however, that defining sets of such syntactic valencies did not characterise words of a given word class in all the positions it could occupy in a sentence. It was concluded that the defining sets of syntactic valencies defined not word classes, but word forms in the morphological paradigm of the word classes. Each word class, in turn, is defined by the distinctive paradigm of morphological forms that have been thus defined. This has the additional advantage of bringing us back to the situation in inflected languages, in which word classes are likewise distinguished by contrasting paradigms of morphological forms.

<sup>28</sup> Another interesting reflection of a culture tradition in the grammar is the subclass of riverine locatives in the nonpersonal nominals (see Mahdi 1993:195), reflecting a riverine orientation of Malay settlements over a considerable part of Indonesian and Malaysian history, partially persisting even till this day.

Use of the same concept of case in a typological environment other than that of inflected languages, to my mind, helps elicit typological particularities of the language. A prominent typological feature of Indonesian grammar is the important role of word order and other word-external means of expressing grammatical categories, as compared with the priority of use of word-internal means in inflected languages. One consequence of this circumstance seems to be that whole groups of words, even relatively long ones, may occupy the syntactic slot of personal nominals. In this, the grammatical feature of 'case form' extends over the entire group, instead of being a characteristic of one word. It is perhaps not a coincidence that English, having lost the greater part of a former Indo-European system of inflection in the paradigms of its nouns, adjectives and verbs, also allows for similar treatment of groups of words. Thus *seven-year-old* behaves as a noun, having the genitive case form *seven-year-old's*, and the plural form *seven-year-olds*.

## References

- Austin, Peter, 1996, Ergativity, clitics and grammatical relations in Sasak. Paper presented at the First Lexical Functional Grammar Conference, Grenoble.
- Cartier, Alice, 1979, De-voiced transitive verb sentences in formal Indonesian. In Frans Plank, ed. *Ergativity: towards a theory of grammatical relations*, 161–183. London: Academic Press.
- Chung, Sandra, 1976, On the subject of two passives in Indonesian. In Charles N. Li, ed. *Subject and topic*, 57–98. New York & London: Academic Press.
- Comrie, Bernard, 1978, Ergativity. In Winfred P. Lehmann, ed. *Syntactic typology*, 329–394. Austin: University of Texas Press.
- Dempwolff, Otto, 1939, *Grammatik der Jabêm-Sprache auf Neuguinea*. Abhandlungen aus dem Gebiet der Auslandskunde, 50. Hamburg: Friederichsen de Gruyter.
- Dixon, Robert M.W., 1972, *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- 1979, Ergativity. *Language* 55:59–138.
- 1994, *Ergativity*. Cambridge Studies in Linguistics 69. Cambridge: Cambridge University Press.
- Du Bois, John W., 1987, The discourse basis of ergativity. *Language* 63:805–855.
- Herrfurth, Hans, 1983, Der abgestufte Ergativcharakter des Satzes als typologisches Merkmal malaio-polynesischer Sprachen und seine valenzgrammatische Erfäßbarkeit. *Acta orientalia* 44:187–207.
- Hopper, Paul, 1983, Ergative, passive and active in Malay narrative. In Flora Klein-Andreu, ed. *Discourse perspectives on syntax*, 67–88. London: Academic Press. (Cited from Du Bois 1987.)
- Lynch, John. 1998, *Pacific languages: an introduction*. Honolulu: University of Hawai'i Press.

- Mahdi, Waruno. 1993, Distinguishing homonymic wordforms in Indonesian. In Ger P. Reesink, ed. *Topics in descriptive Austronesian linguistics*, 181–216. Semaian 11. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië.
- Manning, Christofer D. 1996, *Ergativity: argument structure and grammatical relations*. Dissertations in Linguistics. Stanford: Center for the Study of Language and Information.
- McDonald, R. and Soenjono Dardjowidjojo, 1967, *A student's reference grammar of modern formal Indonesian*. Washington: Georgetown University Press.
- Moeliono, Anton M. and Soenjono Dardjowidjojo, 1988, *Tata bahasa baku Bahasa Indonesia*. Seri Balai Pustaka No. 3657. Jakarta: Balai Pustaka.
- Sneddon, James Neil, 1996, *Indonesian: a comprehensive grammar*. Routledge Grammars. London: Routledge.
- Verhaar, John, W.M., 1978, Some notes on the verbal passive in Indonesian. In John W.M. Verhaar, ed. *Miscellaneous studies in Indonesian and languages in Indonesia* part V, 11–19. NUSA 6. Jakarta: Badan Penyelenggara Seri NUSA.
- Wechsler, Stephen and I Wayan Arka, 1998, Syntactic ergativity in Balinese: an argument against structure based theory. *Natural Language and Linguistic Theory* 16:387–442.
- Zain, St M., 1940, Bahasa Indonesia. *Poedjangga Baroe* 8:16–24.

# Proto Polynesian \*-CIA

---

ANDREW PAWLEY

## 1 Introduction<sup>1</sup>

The thirty or so Polynesian languages form a well-defined subgroup of the large Oceanic branch of Austronesian. In structure and lexicon the subgroup shows a degree of internal diversity roughly comparable to that of the Romance family or the Germanic family. Within Polynesian the primary branching is between a Tongic subgroup (consisting of Tongan and Niuean) and Nuclear Polynesian (all other languages).

Every Polynesian language has a set of suffixes which attach to a verb base, or derive a verb from a noun, and which have the form *-Cia* (where C represents a variable consonant or zero) or, less often, the forms *-a*, *-na* or *-Cina*. *-CIA* is a convenient cover symbol for this set of suffixes. This paper investigates their origins and their functions in Proto Polynesian. The history of the *-CIA* suffixes is perhaps of some general interest as a case of three morphemes (or bits of them) fusing into one as a result of two stages of reanalysis, the first being a shift of morpheme boundary following phonological change and the second a loss of morpheme boundary after one of the constituent suffixes ceased to be fully productive. In some languages subsequent reanalysis of the functions of the now opaque *-CIA* suffixes played a part in restructuring grammatical relations and case marking in verbal clauses.

Individual Polynesian languages generally have between six and eleven alternants of *-CIA*. For example, Maori has eleven: *-a*, *-ia*, *-hia*, *-kia*, *mia*, *-nia*, *-ria*, *-tia*, *-ina*, *-kina* and *-whina*, Samoan has ten: *-a*, *-ia*, *-fia*, *-lia*, *-mia*, *-nia*, *-sia*, *-tia*, *-'ia*, *-ina*, Tongan also has ten: *-a*, *-ia*, *-fia*, *-hia*, *-mia*, *-nia*, *-tia* [sia], *-'ia*, *-ina*, *-kina*, Niuean has nine: *a*, *-ia*, *-hia*, *-kia*, *-mia*, *-nia*, *-tia*, *-ina* and *-na*, and Hawaiian has eight: *-a*, *-ia*, *-hia*, *-kia*, *-lia*, *-mia*, *-nia* and *-na*. In most Polynesian languages it is possible to consider all the *-CIA* suffixes as being

---

<sup>1</sup> It is a pleasure to offer this paper to Byron Bender. In the 1970s I spent several years teaching in the Department of Linguistics at the University of Hawai'i, in the early years of Byron's long tenure as Chair. It was due in no small part to the ambience he created that these years were both extremely pleasant and intellectually stimulating. The first draft of this paper was written in Hawai'i around 1975. Unfortunately, when I came to revise it for publication I was travelling about Europe and had to rely largely on old notes. My inability to consult some relevant sources has left some gaps in my coverage of the evidence, for which I apologise.

variants of a single grammatical marker. The alternants are lexically conditioned. That is to say, a particular base (verb or other) selects a particular alternant but the choice cannot be predicted from the form of the base.

It is a straightforward matter to reconstruct for Proto Polynesian many pairings of a particular base plus a particular alternant of \*-CIA. Table 1 gives a selection of such reconstructed pairings. (The reader will notice that no glosses have been supplied for the suffixed forms in this table. I have omitted these partly because brief glosses can be misleading as to the range of meaning and grammatical functions and partly because I do not want to anticipate the conclusions.)

**Table 1:** Some Proto Polynesian bases that select a specific alternant of -CIA

PPn base		PPn base + suffix
*fono	'join, meet'	*fono-tia
*hala	'way, road; cause'	*hala-nia
*huru	'enter'	*huru-fia
*inu	'drink, drunk'	*inu-mia
*kai	'eat, eaten'	*kai-na
*mafa	'heavy'	*mafa-tia
*matakū	'be afraid'	*matakū-tia
*pulu	'gum (n.); adhere'	*pulu-tia
*quha	'rain (v., n.)'	*quha-ina
*sila	'look, glance'	*sila-fia
*tanu	'bury, buried'	*tanu-mia
*tāji	'cry, weep'	*tāji-sia
*toko	'prop, support, pole (v., n.)'	*toko-nia
*tuqu	'stand (v.)'	*tuqu-ria
*una	'fish scale'	*una-fia

These specific pairings are attested by regular agreements between contemporary Polynesian languages, e.g. PPn \*inu-mia > Maori, Hawaiian, Samoan, Tikopia, Tokelau, Tongan *inumia*; PPn \*tāji-sia > Maori, Tokelau *tājihia*, Hawaiian *kānihia*, Samoan, Tikopia *tājisia*, Tongan *teñihia*; and PPn \*quha-ina > Maori, Samoan, Tokelauan *uaina*, Tikopia *uena*, Tongan *uheina*. The reconstructed pairings are also supported by external evidence, to be discussed in §3. Twelve PPn alternants of \*-CIA can be reconstructed from such widely attested lexical pairings, namely \*-a, \*-fia, \*-kia, \*-lia, \*-mia, \*-nia, \*-ŋia, \*-ria, \*-sia, \*-tia, \*-ina, \*-na.<sup>2</sup> It should be noted that the form \*-ina arose by metathesis from pre Polynesian \*-nia, chiefly when the verb base ended in \*a.

It is more difficult to determine what the functions of \*-CIA were in Proto Polynesian. Among contemporary Polynesian languages -CIA suffixes exhibit a range of functions, chiefly (i)–(vii) below (no single language has the full range):

- (i) marks imperative mood,
- (ii) derives a passive verb from a transitive verb,

<sup>2</sup> It seems the only Proto Polynesian consonants not participating in the \*-CIA suffix were \*p and \*w, and possibly \*h; \*p and \*w continued Proto Oceanic consonants that could not occur word finally.

- (iii) derives a stative verb from a noun or an active intransitive or transitive verb,
- (iv) derives a stative verb from another stative verb with a change of meaning,
- (v) derives a transitive verb that takes ergative case-marking from an intransitive verb or an accusatively marked transitive verb,
- (vi) changes the meaning of an ergative verb,
- (vii) added to an ergative verb has no semantic effect.

### Proto Polynesian case-marking: accusative or ergative?

The question of what function \*-CIA suffixes had in Proto Polynesian has been an important part of a wider debate about Polynesian historical syntax. The broader issue has been whether Proto Polynesian canonical transitive clauses exhibit accusative case-marking, with a productive passive transformation, as is found in Maori and all other well-described Eastern Polynesian languages, or ergative case-marking, as is found in most other Polynesian languages. In the context of this debate the primary function of PPn \*-CIA has been variously perceived as marking a passive derivative, a stative derivative, a transitive derivative or an ergative verb.

Modern debate on this topic was foreshadowed in the 1920s when Herbert Williams (1928) and Spencer Churchward (1928) disagreed about the original uses of PPn \*-CIA. Williams, a specialist in Maori, argued that these uses were best preserved in Maori, where it is generally considered that the main role of -CIA suffixes is to mark passive voice. The active-passive contrast in Maori is exemplified by the typical pair (1a-b):<sup>3</sup>

- (1) a. *Kua inu te wahine i te wai.*  
 PERF drink the woman ACC the water  
 'The woman has drunk the water.'
- b. *Kua inu-mia te wai (e te wahine).*  
 PERF drink-PASS the water by the woman  
 'The water has been drunk (by the woman).'

Such Maori pairs show the typical properties of an active-passive relation. The meaning of the passive verb is predictable, without semantic shifts of the kind common in lexical derivation. In the active construction the Agent NP is obligatory, and unmarked. In the passive the Patient NP is obligatory but the Agent is optional and it is marked by an adposition. However, on closer inspection the Maori passive turns out to have some unusual features, which will be touched on later.

Churchward, a specialist in Samoan, argued that, on the contrary, it is Samoan that more faithfully preserves the original use of \*-CIA as a marker of transitivity—optional in some contexts, obligatory in others. Example (2) shows a pair of Samoan transitive sentences, an affirmative and its negative counterpart. In the affirmative the -CIA suffix -a is optional. In

<sup>3</sup> Key to abbreviations:

ABS – absolutive, ACC – accusative, AG – agent, ART – article, DEF – definite, DUR – durative, ERG – ergative, ES – ergative suffix, INC – inclusive, LOC – locative, n. – noun, N – nominal referent, NP – noun phrase, o.s. – oneself, PASS – passive, PAST – past tense, PERF – perfective aspect, PL – plural, POC – Proto Oceanic, PPn – Proto Polynesian, SG – singular, s.o. – someone, s.th. – something, STAT – stative, TNS – tense, TR – transitive, v. – verb.

the negative, it is obligatory. In Samoan, as in Maori, pre-nominal *e* marks Agent, but as the sentences in (2) are both active, *e* must be interpreted either as a marker of subject or as an ergative marker.

- (2) a. *Sā fasi-(a) le tama e le teine.*  
 PAST hit-CIA the boy ERG the girl  
 'The girl hit the boy.'
- b. *Sā lē fasi-a le tama e le teine.*  
 PAST not hit-CIA the boy ERG the girl  
 'The girl did not hit the boy.'

Following the exchange between Churchward and Williams little was said about these issues until the late 1960s. I will return later to the more recent discussion.

In contemporary Polynesian languages the *-CIA* suffixes are synchronically unanalysable. However, it is known that suffixes of the form *-Cia* are historically an amalgam of elements that once belonged to three separate morphemes: a root-final consonant \*C, a transitive suffix \*-i and a suffix \*-a, whose original function is in dispute.

In the discussion which follows I will first consider the origins of the *-CIA* suffixes—the variation in the initial *-C-*, the *-i-* component, and the *-a* component—and then I will take up the question of what functions the \*-CIA suffixes performed in Proto Polynesian.

## 2 The origin of variable *-C-*

The origins of the lexically conditioned variation in the first element of the *-CIA* suffixes are well known. In Proto Austronesian many word bases or roots ended in a consonant. Many bases were disyllables of the form (C)V(N)CV(C) or trisyllables (C)VCV(N)CV(C), where N represents a nasal whose point of articulation is homorganic with a following obstruent. With one change this state of affairs continued into Proto Oceanic, the immediate ancestor of the Oceanic subgroup, which consists of the Polynesian and Nuclear Micronesian groups and almost all the Austronesian languages of Melanesia. (The change was the reanalysis of Proto Austronesian \*NC clusters as unit phonemes in Proto Oceanic.) In Proto Oceanic, as in Proto Austronesian, many intransitive verbs and nouns ended in a consonant.

After the breakup of Proto Oceanic several daughter languages regularly lost final consonants in absolute word-final position. However, the original final consonant of bases was preserved when a suffix followed. Such suffix-supported stem-final consonants are generally known as 'thematic consonants'.

One of the branches of Oceanic which lost final consonants was that ancestral to Polynesian, Fijian and Rotuman, which together form the Central Pacific subgroup. Proto Polynesian preserved some POC final consonants not only in the \*-CIA forms but also in three other suffixes: \*-Ci, \*Caki and \*-Caŋa. PPn \*-Ci and \*-Caki derived a verb from another verb (with meaning change) or from a noun; PPn \*-Caŋa derived a noun from a verb. Most present-day Polynesian languages retain at least a few reflexes of all four suffixes, although \*-Caki reflexes are no longer fully productive in any contemporary language and reflexes of \*-Ci are fully productive only in Tongan, and then only in a realisation *-i*, which does preserve original thematic consonants. In contemporary languages a particular thematic consonant is often retained in several suffixes. For example, Tokelau *inu* 'drink (v.)' reflects POC \*inum 'drink' with loss of final \*m but retains the thematic consonant \*m both in the



nominalisation *inumanga* ‘beverage’ and in the stative verb *inumia* ‘drinkable, fit for drinking’. Tokelau *tuu* ‘stand (v.)’ loses final \*r from POc \*tuqur ‘stand’ but the thematic consonant is kept (as *l*) in *tuulaki* ‘stand up, get up’, *tuulia* ‘perched on’ and *tuulanga* ‘platform, stand, position, status’. Similarly, final \*s of POc \*tajis ‘cry’ is lost in Maori *tangi* ‘cry, lament’ but kept (as *h*) in the nominalisation *tangihanga* ‘weeping, lamenting, funeral ceremony’ and in the passive *tangihia* ‘wept over, lamented’.

### 3 The origin of -i-

The vowel -i- in the -CIA suffixes continues the Proto Oceanic ‘short’ transitive suffix \*-i (Blust 1986, Evans 2001, Pawley 1973). POc \*-i derived transitive verbs governing a direct object standing in such semantic relations as Patient or Undergoer (of verbs of impact, change of state, etc.), Stimulus (of perception and psychological verbs) and Place (of verbs of posture or bodily excretion). POc \*-i contrasted with the ‘long’ transitivising suffix, \*-akini, which derived transitive verbs with direct objects standing in such semantic relations as Instrument, Concomitant and Cause (Evans 2001, Pawley 1973). Table 2 lists some POc bases which occurred with \*-i.

**Table 2:** Some Proto Oceanic bases and their derivatives  
with the short transitive suffix \*-i

Bases are alphabetised by final consonant, beginning with those that have none.			
base forms		transitive forms	
<i>kani</i>	‘eat’	<i>kani-</i>	‘eat s.th.’
<i>kila</i>	‘know’	<i>kila-i</i>	‘know s.th.’
<i>pitik</i>	‘jerk, flick’	<i>pitik-i</i>	‘jerk or flick s.th.’
<i>totok</i>	‘chop’	<i>tok-i</i>	‘chop s.th.’
<i>tulak</i>	‘push(ed)’	<i>tulak-i</i>	‘push s.th.’
<i>tutuk</i>	‘pounded, smashed’	<i>tuki-</i>	‘pounded, smashed’
<i>gugum</i>	‘clench fist, grasp(ed)’	<i>gumi-</i>	‘grasp s.th.’
<i>inum</i>	‘drink’	<i>inum-i</i>	‘drink s.th.’
<i>tanum</i>	‘earth’	<i>tanum-i</i>	‘bury s.th.’
<i>jalan</i>	‘path, way’ (n.)	<i>jalan-i</i>	‘direct or lead s.o.’
<i>tajan</i>	‘bag’ (n.)	<i>tajan-i</i>	‘bag or net s.th.’
<i>tolon</i>	‘swallow’	<i>tolon-i</i>	‘swallow s.th.’
<i>tokon</i>	‘prop, pole’ (v., n.)	<i>tokon-i</i>	‘support or prop s.th. up’
<i>qucan</i>	‘rain’ (n., v.)	<i>qucan-i</i>	‘rain on s.th.’
<i>quriŋ</i>	‘rudder’ (n.), ‘steer’ (v.)	<i>quriŋ-i</i>	‘steer s.th.’
<i>tatap</i>	‘wash’	<i>tapi-</i>	‘wash s.th.’
<i>qunap</i>	‘fish scale’ (n.)	<i>qunap-i</i>	‘scale a fish’
<i>punuq</i>	‘hit, kill(ed)’	<i>punuq-i</i>	‘hit, kill s.th.’
<i>patuR</i>	‘plait(ed), weave’	<i>patuR-i</i>	‘plait s.th.’
<i>roŋoR</i>	‘hear, heard, be known’	<i>roŋoR -i</i>	‘hear s.th.’
<i>tajis</i>	‘cry’	<i>tajis-i</i>	‘cry for s.th.’
<i>matakut</i>	‘afraid’	<i>matakut-i</i>	‘be afraid of s.th.’

**Table 2:** (cont.) Some Proto Oceanic bases and their derivatives with the short transitive suffix \*-i

<i>mapat</i>	'heavy'	<i>mapat-i</i>	'weigh heavily on s.th.'
<i>bulit</i>	'gum, resin' (n.)	<i>pulit-i</i>	'put gum on, caulk s.th.'
<i>kapit</i>	'carry/carried under the arm'	<i>kapit-i</i>	'carry s.th. under the arm'
<i>kaRat</i>	'bite, bitten'	<i>kaRat-i</i>	'bite s.th.'
<i>kinit</i>	'pinch, nip'	<i>kinit-i</i>	'pinch or nip s.th.'
<i>saqat</i>	'bad'	<i>saqat-i</i>	'dislike s.th.'

Evans (2001) has shown that the Proto Oceanic short transitive suffix had phonologically conditioned variation. When the stem ended in a consonant or \*-a, the suffix appeared as \*-i (e.g. POc \*kila-i-a 'know it' > Fijian (dial.) *kilaya-*). Evans concludes that when the stem ended in a vowel other than \*a (i.e. \*e, \*i, \*o or \*u) the transitive marker \*-i was realised as zero and the stem was immediately followed by an object pronoun, e.g. POc \*kani-a 'eat it' > Fijian *kani-a* 'eat it', PPn \*kai-na 'be eaten, eat!'. It is clear that Pre Oceanic transitive \*-i disappeared (was assimilated) after stem-final \*i or \*e. I am not so sure that that was always the case after \*o or \*u, but these details need not concern us here.

In branches of Oceanic where consonants were lost in absolute word-final position but retained before a suffix, there would have been pattern pressure to shift the morpheme boundary by reassigning final \*C of the base to the suffix, yielding transitive suffixes of the form \*-Ci and \*-Cakini. It is unlikely that such a reanalysis took place immediately after the loss of word-final consonants because at that stage the original final consonants would have been retained before all suffixes, so that the underlying forms of bases would have remained transparent. In time, however, all Oceanic languages which lost final \*C replaced many of the original thematic consonants before suffixes with other consonants. In such languages it is clear that the thematic consonant is no longer regarded as part of the base (Hale 1970, 1973 examines a number of cases where such reanalysis has evidently taken place). In the Polynesian languages, the range of verb stems with no thematic consonant has been extended by the regular loss in all positions of certain POc consonants, such as \*R and (in most languages) \*q.

In some languages, most notably the Fijian group, it can be seen that changes to the original thematic consonants were in part semantically driven. In Fijian, particular consonants became associated with particular types of verb meaning, e.g. in Standard Fijian verbs of motion tend to take *-vi*, verbs of striking *-ki*, verbs of insertion *-mi*, and so on (Arms 1973, 1974). In some Oceanic languages just one or two variants of \*-Ci have become the productive forms, being added to new roots and sometimes competing with or replacing older variants in established verbs.

#### 4 Origins of -a: competing hypotheses

Let us now consider the origins of the vowel -a which, in contemporary languages is the final segment in all variants of PPn \*CIA. Four proposals merit consideration.

- (i) -a comes from a third person singular object pronominal clitic \*-a.
- (ii) -a comes from a stative derivative suffix \*-a, which allows no agent to be expressed.

- (iii) *-a* comes from a passive suffix, which allows an agent to be expressed.
- (iv) *-a* comes from a suffix which derived ergative verbs or was optionally added to ergative verbs.

As a basis for choosing between these alternatives, it will be necessary to compare material from a representative sample of contemporary Polynesian languages and to compare this with cognate material from languages outside the Oceanic group. This in turn will require a discussion of basic clause types.

The hypothesis that *-a* comes from a third person singular object pronominal clitic \*-a was first put forward by Spencer Churchward (1928, 1951) and has been tentatively supported by others. Churchward (1951:74) proposed four stages in the development of the Samoan suffixes *-a*, *-na*, *-Cia* and *-ina* from a prototype exemplified by many languages of Melanesia.

Stage 1. *-a* and *-na* were initially pronominal suffixes. Churchward did not attempt precise formal reconstructions but he was aware that many Austronesian languages of Melanesia use what he called 'anticipatory object pronouns' before direct object nominals (some would say that the object is 'indexed on the verb') and that the third person singular object pronoun in many languages is *-a* and sometimes *-na*.

Subsequent comparative work has confirmed that this pattern of pronominal indexing must be reconstructed for POc (Clark 1973, Evans 2001, Pawley 1973, Ross 1998). When the object of a transitive verb was a specific (as opposed to generic) common noun, the transitive suffix was followed by a suffix marking the person and number of the direct object. The third person singular object pronoun was \*-a, yielding forms such as \*inum-i-a 'drink it', \*matakut-i-a 'be afraid of it/him/her', \*kaRat-i-a 'bite it/him/her'. The verbal morphology of a good many Oceanic languages still conforms to this prototype. Compare the following examples from a Western Fijian language, Wayan:

- (3) a. *Sā sogo tū ne isogo.*  
PERF close DUR ART door  
'The door is closed.'
- b. *Qu sā sogo-ti-a ne isogo.*  
I PERF close-TR-it ART door  
'I have closed the door.'
- c. *Sā lei-sogo-ti ne isogo.*  
PERF PASS-close-TR ART door  
'The door has been closed.'

From sentences (3a, b) it can be seen that the transitive verb *sogo-ti-a* 'close it' is derived from the stative verb *sogo* 'close, be closed' by adding a transitivising suffix *-ti-* followed by an object pronoun. In Wayan and other Fijian languages, many intransitive verb bases are statives and almost all can be transitivised, either by adding *-Ci* or *-Caki*, or by adding a causative prefix plus a transitive suffix. The remaining intransitives are active and these, too, can be transitivised by adding a transitive suffix. Sentence (3c) shows the Wayan pseudo-passive construction, which will become relevant in later discussion. This is derived from the transitive construction by promoting the direct object to subject position, adding a passive (or stativising) prefix *lei-* to the verb and omitting the object pronoun. The transitive suffix is kept. Sentence (3c) differs semantically from (3a) in two ways: it indicates that a change of state has been effected and it implies an Agent.

Stage 2. The suffixes *-a* and *-na* ceased to be pronouns and became just transitive markers. Thus Samoan *fai* ‘do, make’ takes the transitive suffix *-a*, as in ‘*ua na fai-a* ‘he made (it)’, ‘*ua na fai-a le fale* ‘he made (it) the house.

Stage 3. Verbs could be used “without any subject expressed” (Churchward 1951:74), as in *na fasi-a le pua’a* ‘(someone) killed the pig’ and *na fai-a le fale* ‘(someone) built the house, the house was built’. But note that Churchward equates subject with the semantic role of Agent, not with a grammatically defined subject. In these two examples the grammatical subject can only be the Patient.

Stage 4. The ‘subjectless’ constructions were used as true passives. ‘*O le Tusi Pa’ia* ‘*ua fa’asaamoa’ina* ‘The Bible has been translated into Samoan’.

In his pioneering reconstructions of Proto Polynesian syntax Ross Clark (1973, 1976) also tentatively looked towards POc *\*-a* ‘3rd person singular object’ as the source of *\*a-*, with shift of function from object pronoun to simply being part of the transitive suffix (Clark 1973:593). However, I believe there is a much better source. This can be seen clearly if we first look at some Southeast Solomons languages, and then come back to Polynesian, where the evidence is much more complex and confusing.

## 5 Functions of *\*-CIA* and related suffixes

### 5.1 Southeast Solomonic evidence

There is a Southeast Solomonic subgroup of Oceanic centred in Guadalcanal, Gela (Florida), Malaita and Makira (San Cristobal). Evans (2001) has demonstrated the Southeast Solomons languages are among the most conservative daughters of Proto Oceanic in their treatment of transitivity verbal morphology.

It happens that several members of this subgroup also have a detransitivising suffix *-a*. Unfortunately, the material available to me at the time of writing consists mostly of notes from dictionaries of three languages, Arosi, Lau and Gela, which provide only limited information about the semantics and syntax of such derived verbs. However, it is clear that in all the Southeast Solomonic witnesses *-a* is added to transitive verbs carrying the short transitive suffix *-Ci*, that it derives a stative verb (an intransitive verb with Patient subject, usually called a ‘past participle’ in the descriptions) denoting a state resulting from an earlier event and that no Agent can be expressed with the derived stative. In two of the languages, Arosi and Lau, this use of *-a* is productive, occurring in hundreds of forms.

#### Arosi

Arosi, of San Cristobal, belongs to the Cristobal–Malaitan branch of Southeast Solomonic. The suffix *-a* is fairly productive in Arosi. Fox’s (1970) dictionary lists many ‘past participles’ formed by adding *-a* to the short transitive suffix *-Ci* or (less often) to an unsuffixed verb stem. Some of Fox’s glosses are probably misleading in one respect: the bare verb stem is glossed as if the verb were active when it is actually stative (taking a Patient subject and generally implying no agent); the stems *awanga* and *huna* in (4d, g) below are cases in point.

- |     |    |               |                |                |
|-----|----|---------------|----------------|----------------|
| (4) | a. | <i>age</i>    | <i>age-ri</i>  | <i>ageri-a</i> |
|     |    | ‘thatch’ (v.) | ‘thatch s.th.’ | ‘thatched’     |

b.	<i>agu</i> 'climb up, twine around'	<i>agu-ri</i> 'twine around s.th.'	<i>aguri-a</i> 'twined around'
c.	<i>angi</i> 'cry'	<i>angi-si</i> 'cry for s.th.'	<i>angisi-a</i> 'cried for'
d.	<i>awanga</i> 'open up, expose'	<i>awanga-hi</i> 'uncover s.th.'	<i>awangahi-a</i> 'uncovered'
e.	<i>buru</i> 'to stick'	<i>buru-i</i> 'stick s.th. on'	<i>buru-i-a</i> 'caulked'
f.	<i>hau</i> 'plait'	<i>hau-ri</i> 'plait s.th.'	<i>hauri-a</i> 'plaited'
g.	<i>huna</i> 'tie, bind'	<i>huna-si</i> 'tie or bind s.th.'	<i>hunasi-a</i> 'bound'
h.	<i>hunu</i> 'kill'	<i>hunu-i</i> 'kill s.th.'	<i>hunu-i-a</i> 'dead'
i.	<i>mamaa'u</i> 'fear'	<i>mamaa'u-si</i> 'fear s.th.'	<i>mamaa'usi-a</i> 'feared'
j.	<i>ono</i> 'swallow'	<i>ono-mi</i> 'swallow s.th.'	<i>onomili-a</i> 'swallowed'
k.	<i>roku</i> 'fold up'	<i>roku-mi</i> 'fold s.th.'	<i>rokumi-a</i> 'folded up'
l.	<i>siri</i> 'enter'	<i>siri-hi</i> 'enter s.th.'	<i>sirihi-a</i> 'be penetrated, sunk in'
m.	<i>una</i> 'scale of fish' (n.)	<i>una-hi</i> 'scale a fish'	<i>unahi-a</i> 'scaled'
n.	<i>usu</i> 'husk (coconuts)'	<i>usu-ri</i> 'husk s.th.'	<i>usuri-a</i> 'husked'

In Arosi the agent cannot be expressed in clauses whose verb is a derived stative.

### Lau

Lau, a Cristobal–Malaitan language spoken in North Malaita, resembles Arosi in its semiproductive use of *-a* as a stative derivative, added to transitive verbs and to some nouns. The following examples are from the dictionary by Fox (1974).

(5) a.		<i>ade</i> 'make, do'	<i>ade-a</i> 'done'
b.	<i>afe</i> 'in flood, float'	<i>afe-si</i> 'carry s.th. away (current)'	<i>afesi-a</i> 'carried away'
c.	<i>afu</i> 'bear fruit'	<i>afu-si</i> 'beat s.th.'	<i>afusi-a</i> 'fructed plentifully'

d.	<i>angi</i> 'cry'	<i>angi-si</i> 'cry for s.th.'	<i>angisi-a</i> 'cried for, lamented'
e.	<i>fau</i> 'weave'	<i>fau-li</i> 'weave s.th.'	<i>fauli-a</i> 'woven'
f.	<i>ilu</i> 'drink with spoon'	<i>ilu-fi</i> 'drink s.th. with spoon'	<i>ilufi-a</i> 'drunk with a spoon'
g.	<i>kwesu</i> 'alight, to flame'	<i>kwesu-fi</i> 'set s.th. alight'	<i>kwesufi-a</i> 'lit'
h.	<i>laba</i> 'play, trifle'	<i>laba-si</i> 'play with s.th., injure, spoil s.th.'	<i>labasi-a</i> 'injured, spoiled'
i.	<i>laga</i> 'poke'	<i>laga-ni</i> 'poke out joints'	<i>lagani-a</i> 'with joints poked out'
j.	<i>loto</i> 'suck up (liquid)'	<i>loto-fi</i> 'suck s.th. up'	<i>lotofi-a</i> 'soaked'
k.	<i>luba</i> 'draw away, divorce'	<i>luba-si</i> 'divorce s.o.'	<i>lubasi-a</i> 'divorced'
l.	<i>ō</i> 'burnt off (garden)'	<i>ō-fi</i> 'burn s.th. off'	<i>ōfi-a</i> 'burnt off'
m.	<i>'oilaki</i> 'be strong, with spiritual powers'		<i>'oilaki-a</i> 'fortunate, well protected'
n.	<i>'olo</i> 'be straight'	<i>'olo-si</i> 'straighten s.th.'	<i>olosi-a</i> 'straightened'
o.	<i>saru</i> 'burn'	<i>saru-fi</i> 'burn s.th.'	<i>sarufi-a</i> 'burnt'

Lau has a semiproductive use of *-a* to derive stative verbs from nouns (e.g. *rodo* 'night', *rodo-a* 'be dark'; *abu* 'blood', *abu-a* 'be bloody, red').

### Gela

In Gela, a Guadalcanal–Gelic language spoken on Florida Island, *-a* appears to be nonproductive but has similar uses to its Cristobal–Malaitan cognates. Fox (1955) refers to verb forms ending in this suffix as 'past participles' or 'past participle passives'.

(6) a.		<i>ambe</i> 'support s.th.'	<i>ambe-a</i> 'supported'
b.		<i>kari</i> 'scrape s.th. off'	<i>kari-a</i> 'undermined, washed away'
c.	<i>ndura</i> 'pull to pieces'	<i>ndurake</i> 'destroy s.th.'	<i>ndurake-a</i> 'destroyed'
d.		<i>ngiti</i> 'break s.th. in half'	<i>ngiti-a</i> 'be cut into bits'

e.	<i>raranga</i> 'scorching hot'	<i>raranga-si</i> 'dry s.th. up by heat'	<i>rarangasi-a</i> 'dried up, baked by sun'
f.	<i>taligu</i> 'surround, go around'	<i>taligu-ti</i> 'surround, go around'	<i>taliguti-a</i> 'surrounded'
g.	<i>tetevi</i> 'play a stringed instrument'		<i>tetevi-a</i> 'stringed, as a violin'
h.	<i>vane</i> 'grow to abnormal size'	<i>vane-gi</i> 'laden with heavy crop'	<i>vanegi-a</i> 'be laden'
i.	<i>vau</i> 'plait, weave'	<i>vau-hi</i> 'plait, weave s.th.'	<i>vauhi-a</i> 'plaited, woven'
j.		<i>vili</i> 'choose s.th.'	<i>vili-a</i> 'chosen'

In the following sentence, *ahoria* 'tied to the line' derives from the transitive verb *ahori* 'tie (a hook to line)':

- (7) *Na halili t-e ahori-a tua.*  
 the hook TNS-it tie-STAT already  
 'The hook is (already) fixed.'

### General comments

The uses of the stative derivative or 'past participle' suffix *-a* common to Arosi, Lau and Gela can be attributed to Proto Southeast Solomonic, an interstage of Oceanic that is probably somewhat older than Proto Polynesian. It remains uncertain how derived statives ending in *-Cia* differed in meaning from simple stative bases in the Southeast Solomonic languages. A reasonable hypothesis is that simple statives focus on a state of affairs or conditions while statives derived from transitives usually focus more strongly on the earlier action or process from which the state results.

## 5.2 Polynesian evidence

The Southeast Solomonic witnesses agree closely on the functions of *-Cia* verb endings containing the stativiser *-a*. Let us now look at Polynesian languages, whose testimonies are far less consistent. It is convenient to begin with Tongan, because it is morphologically probably the most conservative Polynesian language and because its use of *-CIA* suffixes agrees reasonably well with the Southeast Solomons languages.

### Tongan

Tongan retains *\*-Ci* with a range of thematic consonants only as a nonproductive verb derivative suffix. However, it has a productive suffix *-i* which derives transitive verbs but can also occur in agentless sentences:

- (8) *Na'e tanu-'i 'a e kapa 'e Sione.*  
 PAST bury-TR ABS the can ERG Sione  
 'Sione buried the can.'
- (9) *Na'e tanu-'i 'a e kapa.*  
 PAST bury-TR ABS the can  
 'The can was buried.'

Tongan has a semi-productive suffix *-CIA* with variants *-a*, *-na*, *-ina*, as well as the *-Cia* set which shows a wide range of thematic consonants. C.M. Churchward (1953) calls Tongan *-CIA* a marker of intransitives. Specifically, it seems to derive stative verbs from transitives. Human agents are unacceptable. Compare (10a) and (10b):

- (10) a. *Na'a ku tanu-'i ('a e kapa).*  
 PAST I bury-TR ABS the can  
 I buried it (the can).'
- b. *Na'a ku tanu-mia.*  
 PAST I bury-CIA  
 'I was buried.'
- (11) *Na'e tanu-mia 'a e kapa \*(e Sione).*  
 PAST bury-CIA ABS the can ERG Sione  
 'The can was buried \*(by Sione).'

Chung (1978) points out that some Tongan verbs with *-CIA* mark a completed event. Compare the senses of *tosi* 'nibble' in (12a), a transitive construction, and *tosia* 'successfully nibbled' in (12b), where *tosi-a* is a derived stative verb:

- (12) a. *Tosi 'e he ikā 'a e māta'u.*  
 nibble ERG the fish ABS the hook  
 'The fish nibbles at the hook.'
- b. *'Aho lelei 'eni he na'e tosi-a 'eku māta'u.*  
 day good this because PAST nibble-CIA my hook  
 'It's a good day because my hook has been (successfully) nibbled.'

The *-CIA* verbs sometimes have a durative sense:

- (13) *Na'e hilifaki-a 'a e vakā i he fungahakau.*  
 PAST place-CIA ABS the ship on the reef  
 'The ship was stranded on the reef (for several days).'

Churchward writes that *-CIA* occasionally occurs in transitive constructions. It seems that *-Cia* is sometimes added to a transitive verb when the cause (marked by the ergative 'e) is not a human entity.

- (14) *Na'e fangu-na au 'e he nanamu 'o e kakalā.*  
 PAST awaken-CIA I ERG the smell of the flower  
 'I was awakened by the smell of the flower.'

Tongan also has a suffix *-a* added to nouns to derive stative verbs meaning 'having N, full of N', where N is the referent denoted by the noun. This *-a*, as we will see below, has cognates in Samoan and Fijian.



### Samoan

Samoan retains the POC transitivity marker \*-Ci as a nonproductive suffix, occurring in some hundreds of words. Mosel and Hovdhaugen (1992:205) say that its basic function “seems to be to form words with a more specific and narrow meaning than the words they are derived from”, as in:

- |      |             |              |               |                       |
|------|-------------|--------------|---------------|-----------------------|
| (15) | <i>tala</i> | ‘tell’       | <i>tala-i</i> | ‘summon’              |
|      | <i>ave</i>  | ‘take s.th.’ | <i>ave-i</i>  | ‘be a bearer of news’ |

Affixed to an intransitive verb, -Ci tends to derive a verb, often a transitive verb with a slightly different meaning, for example:

- |      |               |                    |               |   |
|------|---------------|--------------------|---------------|---|
| (16) | <i>gau</i>    | ‘broken’           | <i>gau-i</i>  | ‘break s.th.’                           |
|      | <i>mimita</i> | ‘proud, conceited’ | <i>mita-i</i> | ‘boast (about s.th.)’                   |
|      | <i>nofo</i>   | ‘sit’              | <i>nofo-i</i> | ‘stay, sit down (of a group of people)’ |
|      | <i>o’ono</i>  | ‘strain, grimace’  | <i>ono-si</i> | ‘strain (as a woman in labour)’         |

The derived verb sometimes transforms a simple noun base into a verb denoting a transitive action, one with both agent and patient:

- |      |             |               |                |                      |
|------|-------------|---------------|----------------|----------------------|
| (17) | <i>pulu</i> | ‘rubber, gum’ | <i>pulu-ti</i> | ‘caulk, plug a leak’ |
|      | <i>una</i>  | ‘fish scale’  | <i>una-fi</i>  | ‘scale (a fish)’     |

Chung (1978) and Cook (1978, 1991, 1997) refer to Samoan -CIA as ‘the mysterious suffix’. As well as the -Cia variants, Samoan has -a and -ina. This group of suffixes has a number of fairly distinct functions which grammarians writing on Samoan have found difficult to define precisely. I have referred already to Churchward’s view that these suffixes serve primarily as markers of transitive verbs. Pratt (1911) treats them as passive markers. Milner (1962, 1973) concludes that Samoan -CIA marks perfective aspect (completed events viewed as a whole). While that is often the case, it is not the whole story. Chung (1978) proposes that -CIA has two distinct functions. With middle verbs -CIA is a semiproductive derivational suffix, turning a middle verb into a canonical transitive. In this function it usually has the shape -Cia. With canonical transitives it is a fairly productive ending, and usually has the shape -ina and appears to make little difference to the meaning. Chung suggests the rule: Attach -ina to a canonical transitive if the agent NP is the generic agent, or if the agent has been moved or extracted by a superficial rule.

Mosel and Hovdhaugen (1992:198) divide Samoan -CIA suffixes into several functionally distinct lexical units. They distinguish four morphemes: -a, -ina, -(C)ia and -na. Suffixes with the shape -(C)ia are derivational, usually creating a stative verb from an active verb or a noun. They are not fully productive but occur with scores of verbs. As the examples in (18) indicate, sometimes there is unpredictable semantic change in the derived stative.

(18)	<i>alu</i>	'go, go out'	<i>alu-mia</i>	'be in great demand, sell quickly'
	<i>au</i>	'flow on, continue, reach'	<i>auli-a</i>	'reached, arrived at'
	<i>fuli</i>	'turn over, roll over'	<i>fuli-sia</i>	'turned over'
	<i>masalo</i>	'suspect, think'	<i>masalo-mia</i>	'suspected'

From words referring to weather or other natural phenomena X, *-Cia* forms Patient subject verbs meaning 'be affected by X' (Mosel & Hovdhaugen 1992:203):

(19)	<i>afā</i>	'storm'	<i>afā-tia</i>	'struck by a storm'
	<i>ala</i>	'path, way, method'	<i>alafia</i>	'fair (of wind), suitable'
	<i>ua</i>	'rain' (n., v.)	<i>ua-ina</i>	'rained upon, caught by the rain'
	<i>savili</i>	'breeze, fresh air' (n., v.)	<i>savili-gia</i>	'blown, open to fresh air'
	<i>asu</i>	'smoke' (n., v.)	<i>asu-gia</i>	'affected by smoke'
	<i>sau</i>	'fall (of dew)'	<i>sau-tia</i>	'dewy, dampened'

Some other words, not to do with natural phenomena, also express affectedness:

(20)	<i>ali'i</i>	'chief' (n.)	<i>ali'i-tia</i>	'occupied by chiefs'
	<i>fono</i> (n., v.)	'meeting, hold a meeting'	<i>fono-tia</i>	'serve as the meeting place for a council'
	<i>mālō</i>	'guest'	<i>mālō-ina</i>	'be occupied by guests'
	<i>mata'u</i>	'fear, be afraid'	<i>mata'u-tia</i>	'terrible, dreadful'
	<i>lago</i>	'flies'	<i>lago-ia</i>	'covered with flies'

Mosel and Hovdhaugen (1992:198) say that the four (sets of) forms, *-a*, *-ina*, *-(C)ia* and *-na*, all serve as 'ergativising verbal suffixes', with two main subfunctions:

(i) Changing a nonergative verb into an ergative verb, as in (21):

- (21) a. *Sā alofa Malae iā Vili.*  
 PAST love Malae LOC Vili  
 'Malae loved Vili/Malae felt sorry for Vili.'
- b. *Sā alofa-gia Vili e Malae.*  
 PAST love-ES Vili ERG Malae  
 'Malae cared for Vili/Vili was well-treated by Malae.'

(ii) Making long forms of ergative verbs by adding *-a* or *-ina* to an already ergative verb, often with little or no semantic effect.

- (22) a. *Sā fasi le tama e le teine.*  
 PAST hit the boy ERG the girl  
 'The girl hit the boy.'
- b. *Sā le fasi-a le tama e le teine.*  
 PAST not hit-ES the boy ERG the girl  
 'The girl did not hit the boy.'

Of the various *-CIA* forms that form long forms of ergative verbs, only *-ina* is truly productive. Attached to a nonergative verb *-ina* creates an ergative verb:

- (23) *'Ou ote-gia-ina le tama.*  
 I scold-CIA-ES the boy  
 'I then scolded the boy.'

Samoa has another derivational suffix, *-a*, which derives stative verbs. Mosel and Hovdhaugen (1992: 204–5) label it “ornative” and describe its meaning as “being affected by N”. However, it seems that its primary force is to mark the presence of the referent of N, often an abundant presence. The sense of being affected by N is derived by inference in some cases.

- |      |                 |               |                   |                         |
|------|-----------------|---------------|-------------------|-------------------------|
| (24) | <i>vai</i>      | 'water'       | <i>vai-a</i>      | 'watery'                |
|      | <i>loi</i>      | 'ant'         | <i>loi-a</i>      | 'overrun by ants'       |
|      | <i>namu</i>     | 'mosquito'    | <i>namu-a</i>     | 'full of mosquitoes'    |
|      | <i>'ele'ele</i> | 'dirt, earth' | <i>'ele'ele-a</i> | 'dirty'                 |
|      | <i>fale</i>     | 'house'       | <i>fale-a</i>     | 'provided with a house' |
|      | <i>ma'a</i>     | 'stone'       | <i>ma'a-a</i>     | 'stony'                 |

### Maori

Maori, like other Eastern Polynesian languages, retains reflexes of the short transitive \*-Ci only in a small number of petrified forms. By contrast, *-CIA* suffixes are highly productive. Although Maori *-CIA* suffixes are usually labelled ‘passive’, verbs with this suffix have uses that are not typical of passives. Clark (1973) notes that the ‘passive’ form of the verb is preferred to the active with canonical transitive verbs, that is it is more frequent overall. It is obligatory in imperatives:

- (25) a. *Tua-ina te raakau!*  
 fell-CIA the tree  
 ‘Fell the tree!’
- b. *\*Tua i te raakau.*  
 fell ACC the tree

And it is strongly preferred when the canonical transitive verb is perfective, denoting an event viewed as a complete entity.

- (26) *Ka mau-ria mai e ia ngaa manu.*  
 INC carry-CIA hither AG 3SG DEF.PL bird  
 ‘He brought back the birds.’

Is this preference evidence that, with canonical transitives, the passive construction is not derived from the active but is basic? Chung (1978) gives some arguments against regarding the Maori passive as basic.

### Proto Polynesian case-marking: accusative or ergative?

Discussions of the history of Polynesian syntactic systems have often been complicated by a certain amount of disagreement over the best way to analyse the grammar of particular languages. For example, there has been disagreement over the following questions:

1. What defines a passive construction, in contrast to a stative?
2. What defines the subject of a transitive clause? Indeed, is the notion 'subject' relevant in analysing ergative clauses?
3. What are the basic verb classes of Polynesian languages? Which criteria should be taken as diagnostic?

Biggs (1974) pointed out that the notion 'subject' is problematic in Polynesian languages of the type of Samoan and East Futunan, which are generally regarded as ergative. In such languages the indispensable argument in a transitive clause, and the unmarked argument, is not the Agent but the Undergoer, while the Agent, marked by *e*, is optional. He saw these as reasons to treat the Undergoer NP as the subject and the Agent NP as an oblique case argument. Biggs connected this pattern with the fact that in Polynesian languages there is a basic distinction between two types of intransitive verbs: Actor subject (or active) and Undergoer subject (or stative) verbs. In Samoan and East Futunan, he argued, canonical transitive verbs fall into the Undergoer subject class. Others have preferred the standard ergative analysis, whereby the case marker for subject differs between intransitive and transitive verbs. Still others (Mosel & Hovdhaugen 1992) regard the notion 'subject' as unsuited to the analysis of transitive clauses in Samoan.

In the late 1960s a new generation of scholars reopened the debate that Williams and Churchward had begun 40 years earlier concerning the historical development of transitive clauses in Polynesian.

Following a suggestion made by Kenneth Hale (1968), Patrick Hohepa, a native speaker of Maori, proposed (Hohepa 1969) that Proto Polynesian was essentially like Maori and the other well-described Eastern Polynesian languages, in that it had (i) accusative case-marking, with the subjects of intransitive and active transitive clauses given the same marking, and (ii) a productive passivisation rule. Passives were formed by promoting the direct object to subject position, adding \*-CIA to the verb and marking the Agent with a preposition \**e*. Polynesian languages with ergative case-marking were regarded as reflecting an accusative to ergative 'drift', i.e. the Proto Polynesian system had features which made it prone to reanalysis, with similar reanalyses taking place more than once. Hohepa (1969:314–15) suggested that the Tongan transitive marker *-i* stems from an earlier \*-Cia passive, reflecting a sporadic phonological loss of final vowel which he called 'deletion-from-the right'. The reduced suffixes then lost their passive function. This last part of Hohepa's proposal never gained general acceptance. It is now clear that the verbal suffixes of the form *-Ci* in Tongan, and other Polynesian languages continue the Proto Oceanic transitive suffix \*-i, with the original stem-final consonant reassigned to the suffix.

Hohepa's bold proposals soon drew a response. It was perhaps no coincidence that the first reply came from someone who had recently been doing research on Samoan. Ross Clark (1973, 1976) argued, contrary to Hohepa, that Proto Polynesian was essentially like that found in Samoan, Tongan and Niuean, where transitive constructions that look superficially like Maori passives are usually analysed as ergative constructions, with the direct object of the transitive clause exhibiting the same case-marking as the subject of intransitive verbs. In these languages the Agent of the transitive verb carries a distinct (Ergative) case-marker, preposed *e*, identical in form to the Agent marker of the Maori passive. In ergative languages there is, by definition, no passive.

Clark regarded the distribution of syntactic types across subgroups as important evidence for the historical primacy of the ergative type. It is significant that the ergative pattern is

found not only in both the Tongic languages but also in most Nuclear Polynesian languages, including Samoan, Tokelauan, Futunan and Rennellese. The odd man out is Eastern Polynesian, the subgroup of Nuclear Polynesian that contains Maori and Hawaiian. Clark noted that the two Outlier languages (Nukuoro and Kapingamarangi) which appeared to Hohepa to be accusative in structure turn out, on closer inspection, to be ergative.

Clark concluded that Proto Polynesian resembled Tongan and Samoan in having ergative case marking in canonical transitive clauses. He reconstructed PPn \*-Ci as the unmarked transitive suffix and PPn \*-Cia as a marked transitive suffix. He supposed that \*-Ci was optional, its addition making little difference to the meaning. However, the addition of \*-a to the transitive suffix added “a stative or durative aspectual value” (Clark 1973:589). Clark reached this conclusion on the basis of evidence internal to Polynesian. Tongan *-Cia* and *-a* forms are called “durative” by Churchward, because they show stative aspect. The Tongan *-a* forms also show a preference for agentless constructions and are more polite than *-Ci* verbs. The majority of *-CIA* forms in Niuean are intransitive. However, Clark was unable to find an entirely satisfactory answer to the question of the source of the *-a* in the *-CIA* verbs.

Sandra Chung made several important contributions to the discussion, including Chung (1973, 1978). Although differing with Hohepa on some points she agreed with him on the main issue, that Proto Polynesian had accusative marking and that many Polynesian languages later developed ergative marking. She was not convinced by Clark’s arguments that the Proto Polynesian ergative construction could have been reanalysed as a passive in Eastern Polynesian. Instead she preferred to posit a high frequency passive in Proto Polynesian, like that found in Maori, on the grounds that a high frequency passive yields structural ambiguities that make it prone to reanalysis as an active construction.

### 5.3 Fijian evidence

The immediate relatives of Polynesian are the Fijian languages and Rotuman. Rotuman verb morphology need not detain us; it has been heavily influenced by fairly recent borrowing from Polynesian. That is not the case with the Fijian languages, which in verb morphology can be counted among the more conservative members of the Oceanic group. A consideration of the Fijian evidence provides some clues as to the antiquity of *-CIA* suffixes and their probable functions in Proto Polynesian and earlier stages.

Although they do not agree completely on the morphological details, each of the Fijian languages has a highly productive means of deriving stative verbs (which also function as attributive adjectives) from active verbs and from simple stative verbs. In many descriptions of Fijian languages stative verbs derived by this productive mechanism are referred to as ‘passives’ (Arms 1974, Dixon 1988, Milner 1956), even though the agent is rarely expressed. Indeed, some authors (Milner 1956:97, Schütz 1985:177–179) say that it is impossible to have an agent in Standard Fijian ‘passives’. That is a debatable claim. Speakers of Fijian languages can express certain kinds of instrumental agent, using the locative/instrumental preposition *e* (Standard Fijian) or *i* (Western Fijian and most other dialect areas). They occasionally express a human agent using the dative preposition (*vei* in Standard Fijian, *iva* in Wayan), though this usage may be a calque on English.

Some Fijianists (Schütz & Nawadra 1972, Kikusawa 1998) object to calling such verbs ‘passives’, arguing that they should be classed as derived statives or participles. One can use a range of criteria to define ‘passive’. My view is that, although they seldom occur with an overt human agent, Wayan *lei-* verbs show other features typical of ‘short passives’ (those in

which the agent is absent): (i) a *lei-* verb can be derived from any transitive verb, (ii) the semantic relation between transitive verb and *lei-* verb is predictable, (iii) the *lei-* verb implies a change of state, brought about by an agent, and (iv) the construction is overtly marked in a way that distinguishes it both from constructions with transitive verbs and those with simple stative verbs.

The mechanism used to form passives in Western Fijian languages was briefly described in §4, using an example from Wayan. In Wayan, transitive verbs are typically formed by adding to an intransitive verb either *-Ci* or *-Caki(ni)*, followed by a suffixed or clitic pronoun denoting the direct object, or a proper noun phrase denoting the direct object (e.g. *bola-ti-a* 'cut it open', *buli-ci-a* 'patch it', *butu-ki i Waya* 'set foot on (i.e. visit) Waya'). (*C* can be any of the consonants *c, g, k, l, m, n, r, s, t* or *v*, or zero.) A passive construction is formed by (i) promoting the direct object to subject position, (ii) adding the prefix *lei-* to the transitive form of the verb, and (iii) omitting the direct object pronoun. All or almost all transitive verbs can be passivised in this way. In most passive verb forms the transitive suffix is retained, but in some cases it is omitted and in some it is optional (e.g. *lei-bola* 'be cut open', *lei-butu* or *lei-butu-iki* 'be trodden on, visited').

In most Eastern Fijian languages, including Standard Fijian, there is no passive prefix. A passive verb is formed simply by omitting the direct object pronoun while retaining the transitive suffix, as in Standard Fijian *nanu-mi* 'be remembered, kept in mind', *tagi-ci* 'cried for', *tuku-ni* 'be told', *vakayaga-taki* 'be used'. It can be seen that the 'transitivity' of the *-Ci* and *-Caki* suffixes is semantic rather than grammatical; these suffixes mark the verb as having a semantic role relation to the subject nominal which can be described as 'Object' or 'Patient' (Pawley 1986:88).

Let us now turn to the semantics of passives. Most grammars of Fijian languages have rather little to say on this point. The question arises: How do passive verbs (derived statives) differ, in meaning and discourse use, from simple statives? Arms (1974:51–52) and Schütz (1985:177–179), writing about Standard Fijian, and Dixon (1988:47–48), writing about the closely related Boumaa Fijian language, agree that passives always imply an agent whereas simple statives are indeterminate on the presence or absence of an agent.

The same may be said of Wayan, on which I have worked a fair bit. However, I would prefer to say that Wayan passives (i) mark the state of affairs referred to as being the outcome of a prior action or process, and (ii) imply that this action/process had an agent. 'Agent' does not necessarily imply intent or wilfulness but it does imply a performer who is held responsible. By contrast, a simple stative depicts a state of affairs. It is true that in many cases—depending on the inherent semantics of the particular verb—that state of affairs must be assumed to result from an earlier act or process. However, by choosing the simple stative the speaker backgrounds this consideration. The following are typical sets of Wayan verbs showing (from left to right) the simple stative and the transitive and passive derivatives.

(27) Some simple stative, transitive and passive verbs in Wayan

<i>awa</i> 'far, distant'	<i>awati-a</i> 'keep or put s.th. at a distance'	<i>lei-awati</i> 'kept or put at a distance'
<i>cavu</i> 'extracted, pulled out'	<i>cavuti-a</i> 'extract s.th.'	<i>lei-cavuti</i> 'extracted (by s.o.)'
<i>digi</i> 'chosen, selected'	<i>digi-a</i> 'choose s.th.'	<i>lei-digi</i> 'chosen, selected (by s.o.)'

<i>dokai</i> 'admirable, praiseworthy'	<i>dokaiti-a</i> 'admire or praise s.th.'	<i>lei-dokai</i> 'admired, praised (by s.o.)'
<i>drega</i> 'glued, stuck on'	<i>dregati-a</i> 'glue or stick s.th. on'	<i>lei-dregati</i> 'glued, stuck on (by s.o.)'
<i>kasa</i> 'washed ashore, stranded'	<i>kasavi</i> 'haul or wash s.th. ashore, find s.th. washed ashore'	<i>lei-kasavi-</i> 'washed or hauled ashore, found washed ashore'
<i>kati</i> 'bitten'	<i>kati-a</i> 'bite s.th., nip s.th.'	<i>lei-kati</i> 'bitten, nipped (by s.th.)'
<i>kuya</i> 'be how? in what condition?'	<i>kuyati-a</i> 'do s.th. how?, do in what way?'	<i>lei-kuyati</i> 'done how, occur as a result of what?'
<i>sogo</i> 'close, shut'	<i>sogoti-</i> 'close, shut s.th.'	<i>leisogoti</i> 'closed, shut (by s.o.)'

The semantic contrasts are inadequately expressed by brief dictionary glosses of words cited in isolation. The essential nature of the contrast between simple statives and passives is indicated by the Wayan sentences in (28) and (29).

- (28) a. *Sā sogo tū na sitoa.*  
PERF close PROG ART door  
'The door is closed.'
- b. *Sā lei-sogo-ti tū na sitoa i na ono.*  
PERF PASS-close-TR PROG ART door at ART six  
'The door was closed at six (someone has done it).'
- (29) a. *Sā drega nō na lā ni teveli.*  
PERF glued PROG ART leg of table  
'The table-leg is glued on.'
- b. *Sā lei-drega-ti nō na lā ni teveli.*  
PERF PASS-glued-TR PROG ART leg of table  
'The table-leg is being glued on (by someone).'

I believe that in Standard Fijian the meaning difference between simple statives and passives is typically the same as in Wayan.

Both Eastern and Western Fijian passive verbs differs from the -CIA statives of our Southeast Solomons and Polynesian witnesses in lacking final *-a*. Is this absence an innovation or a retention? The Southeast Solomon and Polynesian evidence suggests that it is an innovation. There is also evidence internal to Fijian which favours the same conclusion.

The Fijian languages also have a number of nonproductive stative deriving affixes. The fullest account of such affixes in Standard Fijian is given by Schütz (1985). There the nonproductive prefixes include the 'spontaneous' stative formatives *ca-*, *ka-*, *ra-*, *ta-*, which generally mark a state as result of an accident or an inanimate cause, and *lau-*, which with human undergoers tends to be associated with adversative meanings (Schütz 1985:218-219) (e.g. *lau-vana* 'shot', *lau-vako* 'nailed'). The nonproductive suffixes in Standard Fijian are *-ca*, *-la*, *-na*, *-ra*, *-ta*, and the one that concerns us here, *-a*. The following selection of pairs,

showing *-a* deriving a stative verb from a noun or, less often, an active verb, is drawn largely from Schütz (1985:221-222). Other Fijian languages show a similar pattern.

(30)	<i>dreke</i>	'cavity'	<i>dreke-a</i>	'(liquid) nearly empty'
	<i>weli</i>	'saliva'	<i>weli-a</i>	'smeared with saliva'
	<i>dravu</i>	'ashes'	<i>dravudravu-a</i>	'covered with ashes, grey'
	(no simple base)		<i>dromodromo-a</i>	'yellow'
	<i>qaro-t-a</i>	'prick it'	<i>qaroqaro-a</i>	'prickly'
	<i>qeke</i>	'earth, soil'	<i>qekeqeke-a</i>	'dirty'
	<i>vere</i>	'plot (v., n.)'	<i>verevere-a</i>	'intricate'
	<i>voto</i>	'thorn'	<i>votovoto-a</i>	'thorny'
	<i>vuti</i>	'body hair'	<i>vutivuti-a</i>	'having much body hair'

This use of *-a* is plainly cognate with the *-a* in Tongan and Samoan which is added to nouns to derive a stative verb denoting the presence or abundance of the referent of the base. It can be seen that most of the Fijian statives with *-a* also show reduplication of the base. Geraghty (1983:265) finds that this pattern of adding *-a* to "make statives out of (usually) reduplicated nominals ... is reconstructable for [Proto Eastern Oceanic] (\*ruavuaravua 'gray' is reflected by Gela *rauravua* and Standard Fijian *dravudravua*)".

Geraghty (1983:263-265) also notes that several derived stative verbs ending in *-Cia* (or dialect variant forms) are widespread within the Fijian group. These forms with *-Cia* differ from the preceding forms in that they are historically derived not from nouns but from verbs. At least two of these derived statives, *caginia* 'blown away (*cagi* 'blow, wind') and *cokotia* (*coko* 'joined, connected, whole'), are found in Western Fijian dialects and in Vanua Levu, Southeast Viti Levu, Kadavu and Lau. Geraghty records several other derived statives ending in *-Cia* although he does not give their geographic distributions: *cevukia* 'blown away' (*cevu* 'float, drift), *dakalia* 'open-mouthed', *kidoria* 'startled' (*kido-ri-* 'startle s.o.'), *katia* 'burned' (*kati-* 'bite, sting or burn s.th.'). These statives look like relic forms, in which stativising *-a* has been added to a transitive verb ending in *-Ci* or *-i*. In fact, *cagi-nia* and *coko-tia* have cognate base forms and suffixes in Polynesian languages: PPN \**añi* 'blow (of wind)', \**añi-na* 'blown away' (showing metathesis from earlier \**añi-nia*) and \**soko* 'join, connected', \**soko-tia* 'gathered together'. Thus, it is possible to reconstruct the Proto Central Pacific pairs \**añi* 'blow', \**añi-nia* 'blown away' and \**soko* 'join, connected', \**soko-tia* 'whole, joined together'. And as we will see in the final section, both of these pairs continue Proto Eastern Oceania etyma.

## 6 Conclusions

The immediate common ancestor of the Southeast Solomonian, Fijian and Polynesian languages, a stage probably very close in time to Proto Oceanic, is usually labelled Proto Eastern Oceanic. A reasonable conclusion from the foregoing comparisons is that Proto Eastern Oceanic had a fairly productive suffix \**-a* which derived stative verbs from transitive verbs and which expressed a state resulting from a prior event. In other words, the derived verbs had the grammatical and semantic force of a 'passive participle' (alias 'perfect



participle’); I will abbreviate this below to ‘passive’. (The passive-deriving \*-a may ultimately have had a common origin with the suffix \*-a which was added to a noun to derive a stative verb denoting the presence or abundance of the nominal referent.)

In Proto Eastern Oceanic the passive derivative \*-a could not be added to verbs carrying the long transitive suffix \*-akini. It could be suffixed only to verbs that carried the short transitive suffix \*-i or to transitive verbs that lacked a transitive suffix. (Short transitives ending in \*-i far outnumbered those without a suffix.) All passivised verbs thus consisted of a verb stem plus the ending \*-ia or \*-a. It is uncertain whether the \*-a passive derivative was fully productive in Proto Eastern Oceanic. While it is evident that many short transitive verbs could take this suffix, it is not clear whether all could. At any rate, from specific lexical agreements between Southeast Solomons, Fijian and Polynesian languages it is possible to reconstruct a number of passive participle forms to the level of Proto Eastern Oceanic. These are given below (third column) together with the corresponding base forms and transitive forms.

(31) a.	<i>aŋin</i> ‘blow (wind)’		<i>aŋin-a</i> ‘blown away’
b.	<i>boŋi</i> ‘be night’	<i>boŋis-i</i>	<i>boŋisi-a</i> ‘be overtaken by night’
c.	<i>bulit</i> ‘stick (v.), gum (n.)’	<i>bulit-i</i> ‘stick s.th. on’	<i>buliti-a</i> ‘caulked’
d.	<i>matakut</i> ‘fear’	<i>matakut-i</i> ‘fear s.th.’	<i>matakuti-a</i> ‘feared’
e.	<i>patuR</i> ‘plait, weave’	<i>patuR-i</i> ‘plait, weave s.th.’	<i>patuRi-a</i> ‘plaited, woven’
f.	<i>sokot</i> ‘be connected, whole’	<i>sokot-i</i> ‘connect s.th.’	<i>sokoti-a</i> ‘solidified, all together’
g.	<i>sulup</i> ‘enter’	<i>sulup-i</i> ‘enter s.th.’	<i>sulupi-a</i> ‘be penetrated, sunk in’
h.	<i>taŋis</i> ‘cry’	<i>taŋis-i</i> ‘cry for s.th.’	<i>taŋisi-a</i> ‘cried for, lamented’
i.	<i>qusan</i> ‘rain’ ‘rain’	<i>qusan-i</i> ‘rain on s.th.’	<i>qusani-a</i> ‘rained on’
j.	<i>unap</i> ‘scale of fish’ (n.)	<i>unap-i</i> ‘scale a fish’	<i>unapi-a</i> ‘scaled’

As noted earlier, stem-final consonants were lost in word-final position in the Southeast Solomonic and Central Pacific branches of Eastern Oceanic. Word internally they were reanalysed as part of the suffix, yielding a reanalysis of the short transitive suffix as \*-Ci, with their derived passives now ending in \*-Cia.

Twenty-odd years ago I drew Sandra Chung’s attention to the Southeast Solomons evidence and suggested that these languages had preserved an old, Proto Eastern Oceanic construction, in which \*-a was added to transitive \*-Ci to derive a stative verb or adjective. Chung (1978:258) comments that: “If \*-a can indeed be reconstructed as an adjectival suffix in Proto Eastern Oceanic, and \*-Cia as the suffixal morphology of passive adjectives derived

from verbs, then the latter would seem closer to Hale-Hohepa's passive suffix than to Clark's transitive suffix".

That is true. Many Proto Eastern Oceanic passive participles were retained in Proto Polynesian, as verbs ending in \*-Cia, \*-a or \*-ina. However, we need not conclude that these agreements between Southeast Solomons and Polynesian languages strongly support the hypothesis that Proto Polynesian had accusative marking, with a full passive and productive formation of passive verbs by the suffix \*-CIA. I believe Clark's reconstruction of Proto Polynesian case-marking as ergative is probably correct. My main point of disagreement is with his proposal that PPn \*-Cia was a variant or extension of the \*-Ci transitive suffix, attached to ergative verbs, and that the most likely source of \*-a was POc \*-a '3rd person singular object' (Clark 1973:593). I have tried to show in this paper that there is a better source.

I think the balance of the internal Polynesian evidence favours the view that PPn \*-CIA was a fairly productive suffix. It could be added to some transitive verbs but not all. As to whether \*-CIA verbs could take an agent, the Polynesian evidence is mixed. In Tongan, human agents are absent or extremely rare with -CIA verbs. The external evidence is consistent with Tongan. Human agents are not allowed with passive verbs in the Southeast Solomon languages and their rare occurrences in Fijian discourse may well be calques on English. Thus, the chances are that, in Proto Eastern Oceanic and Proto Central Pacific, constructions headed by a passive participle were agentless.

But how the case-marking system reconstructable for Proto Eastern Oceanic changed into the types found in Tongan, Samoan or the Eastern Polynesian languages is a question too large to tackle here. Polynesianists have debated it for some time and no doubt will continue to do so.

## References

- Arms, David, 1973, Whence the Fijian transitive endings? *Oceanic Linguistics* 12:503–558.
- 1974, Transitivity in Standard Fijian. PhD thesis, Department of Linguistics, University of Michigan, Ann Arbor.
- Biggs, Bruce, 1974, Some problems of Polynesian grammar. *Journal of the Polynesian Society* 83:4:401–426.
- Blust, Robert A., 1986, A rediscovered Austronesian comparative paradigm. *Oceanic Linguistics* 16:1–51.
- Chung, Sandra, 1973, On the syntax of nominalizations in Polynesian. *Oceanic Linguistics* 12:641–686.
- 1978, *Case-marking and grammatical relations in Polynesian*. Austin & London: University of Texas.
- Churchward, C.M., 1953, *Tongan grammar*. Oxford: Oxford University Press.
- Churchward, Spencer, 1928, On the origins of the Polynesian passive. *Journal of the Polynesian Society* 37:300–305.
- 1951, *A Samoan grammar* (2nd edn). Melbourne: Spectator.

- Clark, Ross, 1973, Transitivity and case in Eastern Oceanic languages. *Oceanic Linguistics* 12:559–605.
- 1976, *Aspects of Proto-Polynesian syntax*. Auckland, N.Z.: Linguistic Society of New Zealand.
- Cook, Kenneth W., 1978, The mysterious Samoan transitive suffix. *Berkeley Linguistic Society* 4.
- 1991, The Samoan *cia* suffix as an indicator of agent defocusing. *Pragmatics* 1(2):145–167.
- 1997, The Samoan transitive suffix as an inverse marker. In M. Verspoor, Kee Dong Lee and E. Sweetser, eds *Lexical and syntactical constructions and the construction of meaning*, 347–361. Amsterdam/Philadelphia: Benjamins.
- Dixon, Robert M.W., 1988, *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- Evans, Bethwyn, 2001, A study of valency-changing devices in Proto Oceanic. PhD thesis, Department of Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Fox, Charles E., 1955, *A dictionary of the Nggela language (Florida, British Solomon Islands)*. Auckland: Unity Press.
- 1970, *Arosi dictionary*. Canberra: Pacific Linguistics.
- 1974, *Lau dictionary*. Canberra: Pacific Linguistics.
- Geraghty, Paul, 1983, *The history of the Fijian languages*. Oceanic Linguistics Special Publication No. 19. Honolulu: University of Hawai'i Press.
- Hale, Kenneth W., 1968, Review of Hohepa 1967, A profile-generative grammar of Maori. *Journal of the Polynesian Society* 77:83–99.
- 1970, The passive and ergative in language change: the Australian case. In S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic studies in honour of Arthur Capell*, 757–781. Canberra: Pacific Linguistics.
- 1973, Deep-surface canonical disparities in relation to analysis and change: an Australian example. In T.E. Sebeok, ed. *Current trends in linguistics* Vol. 11, 401–58. The Hague: Mouton.
- Hohepa, Patrick W., 1969, The accusative-to-ergative drift in Polynesian languages. *Journal of the Polynesian Society* 78:295–329.
- Kikusawa, Ritsuko, 1998, A formal analysis of the so-called 'passive' in Fijian. *Journal of Asian and African Studies* 56:113–140.
- Milner, George B., 1956, *Fijian grammar*. Suva: Government Press.
- 1962, Active, passive or perfective in Samoan: a fresh appraisal of the problem. *Journal of the Polynesian Society* 71:151–161.
- 1973, It is aspect (not voice) that is marked in Samoan. *Oceanic Linguistics* 12:621–639.
- Mosel, Ulrike and Even Hovdhaugen, 1992, *A reference grammar of Samoan*. Oslo: Scandinavian University Press/Institute for Comparative Research in Human Culture.

Pawley, Andrew, 1973, Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12: 103–88.

1986, A reanalysis of Fijian transitive constructions. In Chris Corne and Andrew Pawley, eds *Le coq et le cagou: essays on French and Pacific Languages*, 81–112. Auckland: Linguistic Society of New Zealand.

Pratt, George, 1911, *Grammar and dictionary of the Samoan language* (4th edn). Malua, Western Samoa: Malua Printing Press.

Ross, Malcolm, 1998, Proto Oceanic phonology and morphology. In M. Ross, A. Pawley and M. Osmond, eds *The lexicon of Proto Oceanic*. vol. 1: *Material culture*, 14–35. Canberra: Pacific Linguistics.

Schütz, Albert J., 1985, *The Fijian language*. Honolulu: University of Hawaii Press.

Schütz, Albert J. and Tevita Nawadra, 1972, A refutation of the notion “passive” in Fijian. *Oceanic Linguistics* 11(2):88–109.

Williams, H., 1928, Some observations on Polynesian verbs. *Journal of the Polynesian Society* 37:306–317.

# *Pohnpeian possessive paradigms: the smart solution, the dumb solution and the Pohnpeian solution*

---

KENNETH L. REHG

This paper evaluates three possible analyses of the alternations in vowel qualities that occur when possessive pronouns are suffixed to nominal bases in Pohnpeian direct possessive paradigms. The first two approaches are labelled ‘the smart solution’ and ‘the dumb solution’, in imitation of Hale’s (1973) discussion of transitive suffixes in Maori. Neither one proves satisfactory. Instead, the superior analysis appears to be ‘the Pohnpeian solution’, which compromises one of the desiderata of generative phonology—that there be a single underlying representation for each morpheme—but accounts very well for the complex patterns of vowel alternations in Pohnpeian.

## 1 Prelude<sup>1</sup>

The earliest version of this paper was presented in the spring of 1982 as a talk given to the Austronesian Circle in Honolulu, Hawai‘i. That presentation was accompanied by an extensive handout that was circulated among my colleagues, and references to it sometimes occur in the literature on Micronesian languages. In 1986, I reworked that material and incorporated it into chapter 6 of my dissertation (Rehg 1986); the chair of my dissertation committee was Byron W. Bender, to whom this volume is dedicated.

The present version of this paper is a revised and, I hope, improved rendering of that earlier, unpublished work. It continues, however, to focus on a concern that was common among some of the Micronesians at the University of Hawai‘i in the 1970s and early 1980s:

---

<sup>1</sup> I am deeply indebted to Rodrigo Mauricio, who searched the *Ponapean-English dictionary* (Rehg & Sohl 1979) for all directly possessed nouns and organised them according to paradigm types. I also wish to acknowledge Damian Sohl and Marcelino Actouka, who spent a considerable amount of time discussing these data with me, as well as Joel Bradshaw, who made many useful suggestions for improving this work. Robert Andreas assisted me by providing additional data, and Kimi Miyagi discussed the Japanese forms with me. The shortcomings of the analysis are, of course, my responsibility alone.

that one of the imperatives of generative phonology—that allomorphy be minimised—is sometimes at odds with the data, typically in very subtle ways. Those of us who shared this view read Hale (1973) with great interest, and this paper clearly reflects the influence of his insightful work. It also exemplifies the considerable impact that Byron has had on my research activities. Indeed, much of what I have written in the past has been inspired by his pioneering work on the languages of Micronesia.<sup>2</sup>

## 2 Introduction

In Pohnpeian,<sup>3</sup> as in most Oceanic languages, possessive constructions are of two types. One type entails **direct possession**, where the morpheme(s) identifying the possessor follow the noun representing the possessed; the second type involves **indirect possession**, where the opposite order occurs.<sup>4</sup> Examples are given in (1):<sup>5</sup>

- |     |              |                              |
|-----|--------------|------------------------------|
| (1) | <i>mɔŋey</i> | <i>nimey u:p<sup>w</sup></i> |
|     | head-my      | CLS-my drinking coconut      |
|     | 'my head'    | 'my drinking coconut'        |

An extensive discussion of Pohnpeian possessive constructions is provided in §4.8 of the *Ponapean reference grammar* (Rehg & Sohl 1981). The focus in this paper is specifically on direct possession and the alternations in vowel qualities that occur when possessive pronouns are suffixed to nominal bases. A sample paradigm given in phonemic transcription appears in (2). Note in the Pohnpeian forms that the lower-mid front vowel *ɛ*, which occurs in the singular and construct forms, alternates with the low vowel *a*, which occurs in the free and nonsingular forms.

- |     |            |                         |  |
|-----|------------|-------------------------|--|
| (2) | Free Form  | <i>ma:s</i>             | 'face'                                 |
|     | 1SG        | <i>meseɣ</i>            | 'my face'                              |
|     | 2SG        | <i>mese<sup>m</sup></i> | 'your face'                            |
|     | 3SG        | <i>mese</i>             | 'his/her/its face'                     |
|     | IDU/PL EXC | <i>masat'</i>           | 'our (exclusive) face(s)' <sup>6</sup> |
|     | IDU INC    | <i>masat'a</i>          | 'our (inclusive) face(s)'              |

<sup>2</sup> For example, see Bender 1973 for a discussion of some of the phenomena considered in this paper.

<sup>3</sup> The name of the island where this language is spoken was changed from Ponape to Pohnpei in 1981; the use of **Pohnpeian** as the preferred English name for the language is more recent.

<sup>4</sup> In the literature on Oceanic languages, **direct** and **indirect** patterns of possession are traditionally labelled **inalienable** and **alienable**, respectively. Here I follow the usage in the *Ponapean reference grammar* (Rehg & Sohl 1981), where the terms **direct** and **indirect** were first used.

<sup>5</sup> This paper employs conventional phonetic symbols to represent the phonemes of the northern dialect of Pohnpeian, except that *p<sup>w</sup>* and *m<sup>w</sup>* are labials that are both rounded and velarised, while *t'* is a voiceless laminal alveolar stop (previously described as a voiceless retroflexed affricate). See Rehg and Sohl 1981 for a discussion of the phonemic inventory of Pohnpeian. Abbreviations are: 1 – first person, 2 – second person, 3 – third person; in example (1) CLS – classifier (there are more than 20 possessive classifiers in Pohnpeian); C – consonant, DU – dual, EXC – exclusive, INC – inclusive, PL – plural, SG – singular, V – vowel; in derivations, SR – surface representation and UR – underlying representation.

<sup>6</sup> 'Our (exclusive) faces' is, in fact, *masat'akat'*, where *akat'* is an enclitic marking plurality; hence, I enclose *s* in the English translations within parentheses.

2DU	<i>masam<sup>w</sup>a</i>	'your face(s)'
3DU	<i>masara</i>	'their face(s)'
1PL INC	<i>masat' ayl</i>	'our (inclusive) face(s)'
2PL	<i>masam<sup>w</sup> ayl</i>	'your face(s)'
3PL	<i>masarayl</i>	'their face(s)'
Construct Form	<i>mesen</i>	'face of'

As additional data will illustrate, accounting for the alternations between these two vowels is by no means a straightforward task. Three possible analyses will be considered here. The first two of these will be labelled 'the smart solution' and 'the dumb solution', in imitation of Hale's discussion of transitive suffixes in Maori (Hale 1973). A third solution, called 'the Pohnpeian solution', will then be advanced as the preferable analysis of the data.

### 3 The smart solution

Within standard generative phonological theory, a **smart** solution is one in which each morpheme has a single underlying representation, the surface forms of which are derived by a set of ordered rules. A first, ultimately unsuccessful, approximation of such a smart solution—one that I formulated in the early 1970s within *The Sound Pattern of English* framework (Chomsky & Halle 1968)—was essentially like the following, which attempted to account for forms like those for 'face' listed in (2).<sup>7</sup> In this solution, all nominal bases were assumed to end in vowels, consistent with the historical facts of the language.

(3) BASE FORMS<sup>8</sup>

'face'	<i>masa</i>				
1SG	<i>-yi</i>	1nonSG	<i>-t'<sup>9</sup></i>	DU	<i>-aa</i>
2SG	<i>-m<sup>w</sup>i</i>	2nonSG	<i>-m<sup>w</sup></i>	PL	<i>-ayl</i>
3SG	<i>-i</i>	3nonSG	<i>-r</i>	Construct	<i>-ni</i>

(4) Ordered Rules (informally stated)

1. Nominal Lengthening

$[\#(C)V(C)(V)\#]_N \Rightarrow [\#(C)V:(C)(V)\#]_N$

This is an early 1970s statement of the rule of nominal lengthening discussed in detail in Rehg 1984 and in §5.3 of Rehg 1986.

<sup>7</sup> This early analysis was modelled after one proposed by Irwin Jay Howard (pers. comm.) for similar data in Trukese (also known as Chuukese). Other variants of this solution can be found in my early studies of this language.

<sup>8</sup> One might posit *-i* as a singular marker, but I did not do so in my early analyses of Pohnpeian.

<sup>9</sup> The 1EXC form does not combine with either dual or plural suffixes; it is simply nonsingular. That is, it can be used with two or more referents.

2. Low Vowel Raising<sup>10</sup>

$$\left[ \begin{array}{c} a \\ \text{--long} \end{array} \right] \rightarrow \epsilon / \_ C_0 i \#$$

This rule raises short /a/ to /ε/ in partial assimilation to /i/, regardless of the number of intervening consonants.

## 3. Vowel Copying

$$\left[ \begin{array}{c} a \\ \text{--long} \end{array} \right] \rightarrow \epsilon / \_ C_0 \epsilon$$

This rule states that short /a/ will raise to /ε/ if /ε/ occurs in a following syllable, regardless of the number of intervening consonants.

## 4. Final Vowel Deletion

$$V \rightarrow \emptyset / V(C) \_ \#$$

Given that in this analysis I treated long vowels as underlying  $V_i V_i$  sequences, this rule has the effect of shortening final long vowels and deleting final short vowels.<sup>11</sup>

These rules, when applied to the base forms above, generate the correct surface forms for the paradigm previously presented. Sample derivations for the free form of 'face', as well as for the singular and construct forms, are shown in (5):

(5) UR	<i>masa</i>	<i>masa+yi</i>	<i>masa+m<sup>w</sup>i</i>	<i>masa+i</i>	<i>masa+ni</i>
1. Lengthening	<i>ma:sa</i>	--	--	--	--
2. Raising	--	<i>mase+yi</i>	<i>mase+m<sup>w</sup>i</i>	<i>mase+i</i>	<i>mase+ni</i>
3. Copying	--	<i>mese+yi</i>	<i>mese+m<sup>w</sup>i</i>	<i>mese+i</i>	<i>mese+ni</i>
4. Final V Deletion	<i>ma:s</i>	<i>mese+y</i>	<i>mese+m<sup>w</sup></i>	<i>mese</i>	<i>mese+n</i>
SR	<i>ma:s</i>	<i>mesey</i>	<i>mese<sup>w</sup></i>	<i>mese</i>	<i>mesen</i>

Note that, in this analysis, Rule 2 is crucially ordered before Rule 3, so as to provide the input for copying. Rule 2 must also be ordered before Rule 4; therefore, Rule 2 requires the presence of word-final *i*. In this early attempt to account for these data, Rule 1 was also ordered before Rule 2, to account for data like (6). Therefore, it was assumed that Lengthening had to apply prior to Raising so that raising would be blocked. No crucial evidence for ordering between Rules 3 and 4 was established.

(6) BASE FORM	GLOSS	SURFACE FORM
<i>laŋi</i>	'sky'	<i>la:ŋ</i>
<i>sali</i>	'rope'	<i>sa:l</i>
<i>aŋi</i>	'wind'	<i>a:ŋ</i>

Two of the rules cited above—Lengthening and Final Vowel Deletion—have been discussed at length in my dissertation and elsewhere (e.g. Rehg & Sohl 1981; Rehg 1984, 1993); these rules are motivated by forms other than just those discussed here, and they persist in my analysis of Pohnpeian, albeit in different forms. The rules of Raising and Vowel Copying were introduced in this analysis to account for alternations between *a* and *ε*. They

<sup>10</sup> I employ the term 'raising' throughout this paper when talking about *a* becoming *ε*. In part, this is because I treat *a* as [-back] in this paper, for reasons discussed in Rehg 1986. There is clearly a substantial difference in jaw height between *a* and *ε* in Pohnpeian.

<sup>11</sup> In my earlier work, written within a linear framework, I treated underlying long vowels as a sequence of two identical vowels and derived/surface long vowels as a single vowel plus a feature of length. This strategy was employed in an attempt to account for the Janus-faced nature of long vowels, now dealt with more effectively within the framework on nonlinear phonology.



correctly account for the data considered thus far, as well as for other data not yet considered. For example, note that both rules are constrained to apply to short *a*; neither raising nor copying affects long vowels, as illustrated in the paradigms in (7), where underlying long vowels are represented as a sequence of two identical vowels and derived/surface long vowels are represented as /V:/.

(7) Base Forms	<i>nt'aa</i>	'blood' <sup>12</sup>	<i>kaawa</i>	'buttocks'
1SG	<i>nt'a:y</i>		<i>ka:wey</i>	
2SG	<i>nt'a:m<sup>w</sup></i>		<i>ka:wem<sup>w</sup></i>	
3SG	<i>nt'a:</i>		<i>ka:we</i>	
1DU/PL EXC	<i>nt'a:t'</i>		<i>ka:wat'</i>	
1DU INC	<i>nt'a:t'a</i>		<i>ka:wat'a</i>	
2DU	<i>nt'a:m<sup>w</sup>a</i>		<i>ka:wam<sup>w</sup>a</i>	
3DU	<i>nt'a:ra</i>		<i>ka:wara</i>	
1PL INC	<i>nt'a:t'ayl</i>		<i>ka:wat'ayl</i>	
2PL	<i>nt'a:m<sup>w</sup>ayl</i>		<i>ka:wam<sup>w</sup>ayl</i>	
3PL	<i>nt'a:rayl</i>		<i>ka:warayl</i>	
Construct	<i>nt'a:n</i>		<i>ka:wen</i>	
Free Forms	<i>nt'a</i>		<i>ka:w</i>	

The requirement that the high vowel in the environment of the rule of Raising be before a word boundary—that is, in word-final position—was motivated by the fact that /a/ fails to raise to /ε/ in forms like those in (8):

(8) <i>lakit</i>	'to discard'
<i>m<sup>w</sup>atik</i>	'trickery'
<i>paŋin</i>	'to awaken'

The rule of Copying also gains support from the data in (9), provided that this rule is allowed to apply iteratively, right to left, across the string.

(9) Base Form	Gloss	Free Form	Construct Form
<i>likaraka</i>	'louse'	<i>likarak</i>	<i>likereken</i>
<i>wakara</i>	'pubic hair'	<i>wakar</i>	<i>wekeren</i>
<i>kaparapara</i>	'fertile'	<i>kaparapar</i>	<i>kepereperen</i>

Other data that were not considered in the formulation of this analysis, however, reveal that these rules cannot account for all the relevant facts concerning alternations between *a* and *e*. Consider therefore (10), where the base forms 'sky' and 'wind' are followed by the construct suffix *-ni* and the adjectival suffix *-na*. Note in these examples that raising is triggered by underlying /-ni/ (but not by /-na/), even though a syllable of the shape /Ci/ intervenes. That is, raising occurs even though a high vowel intervenes between *a* (the vowel subject to raising) and word-final *i* (the vowel triggering the raising).

(10) Base Form	Construct Form	Adjectival Form	Free Form	Gloss
<i>laŋi</i>	<i>leŋin</i>	<i>laŋin</i>	<i>la:ŋ</i>	'sky'
<i>aŋi</i>	<i>eŋin</i>	<i>aŋin</i>	<i>a:ŋ</i>	'wind'

<sup>12</sup> The initial nasal in the word for 'blood' is syllabic on the surface. See §2.9.2 of *Ponapean reference grammar* (Reh & Sohl 1981) for further discussion.

I once thought it might be possible to salvage the rule of Raising by reformulating it as (11). However, even after modifying the rules of Copying and Raising, it is still impossible to account for the full range of alternations between *a* and *ε*. Compare, for example, the paradigms in (12):

$$(11) \left[ \begin{array}{l} a \\ -\text{long} \end{array} \right] \rightarrow \epsilon / \_ (C_0 i)_1 \#$$

(12) Base Forms	<i>wara</i>	'canoe' <sup>13</sup>	<i>m<sup>w</sup>are</i>	'title'
1SG	<i>werey</i>		<i>m<sup>w</sup>arey</i>	
2SG	<i>werem<sup>w</sup></i>		<i>m<sup>w</sup>arem<sup>w</sup></i>	
3SG	<i>were</i>		<i>m<sup>w</sup>are</i>	
1DU/PL EXC	<i>warat'</i>		<i>m<sup>w</sup>aret'</i>	
1DU INC	<i>warat'a</i>		<i>m<sup>w</sup>arat'a</i>	
2DU	<i>waram<sup>w</sup>a</i>		<i>m<sup>w</sup>aram<sup>w</sup>a</i>	
3DU	<i>warara</i>		<i>m<sup>w</sup>arara</i>	
1PL INC	<i>warat'ayl</i>		<i>m<sup>w</sup>arat'ayl</i>	
2PL	<i>waram<sup>w</sup>ayl</i>		<i>m<sup>w</sup>aram<sup>w</sup>ayl</i>	
3PL	<i>wararayl</i>		<i>m<sup>w</sup>ararayl</i>	
Construct	<i>weren</i>		<i>m<sup>w</sup>aren</i>	
Free Forms	<i>wa:r</i>		<i>m<sup>w</sup>a:r</i>	

Note that while all instances of /a/ raise to /ε/ in the singular and construct forms in the paradigm for 'canoe' (as expected), the paradigm for 'title' cannot be explained; in fact, it illustrates that the previously formulated Copying rule is incorrect. That is, the presence of /ε/ in the second syllable of the singular, first person dual/plural exclusive and construct forms in the 'title' paradigm does not trigger the raising of preceding /a/. Underlying /ε/, in fact, **blocks** the application of Copying, as further illustrated by (13). The lowering of /ε/ that occurs in the dual and plural forms of the paradigm for 'title' is also not accounted for by the rules thus far presented, but this phenomenon will not be further considered until §5 of this paper.

(13) Base Form	Gloss	Construct Form
<i>kat'εpa</i>	'worth'	<i>kat'εpen</i>
<i>apera</i>	'shoulder'	<i>aperen</i>
<i>saleŋa</i>	'ear'	<i>saleŋen</i>

In light of these data, and others not yet considered, it eventually became apparent that the alternations between *a* and *ε* that occur in Pohnpeian possessive constructions could not readily be accounted for within the theoretical framework of *The Sound Pattern of English*. However, work initiated by Paul Kiparsky (1968), on what he later came to call the 'Non-derived Environment Blocking syndrome' (Kiparsky 1993), appeared to provide a way to deal with these difficult data. An early, straightforward account of Non-derived Environment Blocking was provided by Kiparsky (1982, cited here from Goldsmith 1999:49) as his "Revised Alternation Condition".

<sup>13</sup> When suffixed by possessive pronouns, the base *wara* functions as a possessive classifier; the phonological rules governing the forms of possessive classifiers and directly possessed nouns, however, are identical. See §4.8 of the PRG for further discussion.

(14) **Revised Alternation Condition (RAC)**

Obligatory neutralization rules apply only in derived environments.

Definition: An environment E is **derived** with respect to a rule R if E satisfies the structural description of R crucially by virtue of a combination of morphemes or the application of a rule.

“That is”, Kiparsky explains, “an obligatory neutralization rule can apply only if the input involves crucially a sequence which arises in morpheme combinations or through the earlier application of a phonological rule”.

The theoretical basis of this syndrome has been the subject of considerable debate (Mascaró 1976; Kiparsky 1982, 1993; Iverson & Wheeler 1988; Cole 1995, etc.), but I will not further explore these controversies here. What is important for the purposes of this paper is that, if one incorporates this observation into one’s grammar, by whatever means, the phonological analysis of Pohnpeian directly possessed nouns can be considerably simplified.

First, given the RAC, it is no longer necessary to order Lengthening before Raising, as in the derivation in (5), since the raising of *a* to *ɛ* is an obligatory neutralisation rule (*a* and *ɛ* are both phonemes in Pohnpeian). Therefore, Raising cannot apply in a monomorphemic form like *lanji* ‘sky’, but it can apply in polymorphemic forms; hence, /*lanji+ni*/ becomes [lɛnjin].<sup>14</sup>

Second, the RAC permits a simpler account of the rule of Low Vowel Raising, which may be informally reformulated as in (15):

$$(15) \text{ Low Vowel Raising} \\ \left[ \begin{array}{c} a \\ \text{-long} \end{array} \right] \rightarrow \epsilon / \_ (C_0i)C_0i^{15}$$

The rule of Vowel Copying is also revised here, as in (16), for reasons to be considered below.

$$(16) \text{ Vowel Copying} \\ \left[ \begin{array}{c} a \\ \text{-long} \end{array} \right] \rightarrow \epsilon / \_ (C_0i)C_0\epsilon^{16}$$

Although it might seem obvious that rules (15) and (16) should be collapsed into a single rule, for the purposes of this section I will keep them separate. I will, however, return to this issue in §5.

Note that, as a consequence of the RAC, it is no longer necessary to include boundaries in (15).<sup>17</sup> Given a string of the shape /*wara+ni*/ ‘canoe of’, which involves a combination of

<sup>14</sup> The rule of Lengthening can also be simplified so that it applies only to monomoraic nouns, subsequent to Final Vowel Deletion.

<sup>15</sup> This rule, if correctly formulated, would appear to be problematic for at least some versions of radical underspecification theory; /i/ serves as both a trigger and as a transparent segment. It is not clear how such a rule could be treated in accord with the Redundancy Rule Ordering Constraint proposed by Abaglo and Archangeli (1989:474). Steriade (1995:129-130) notes a similar problem for Russian.

<sup>16</sup> Both (15) and (16) might be expressed within the framework of autosegmental phonology as feature spreading rules. However, because this issue does not bear on the argument being developed here, and because the audience for this book includes morphologists as well as phonologists, I employ more traditional notational devices.

<sup>17</sup> As noted in §5, however, a ‘domain’ must be specified for this rule.

morphemes, the rule simply applies right to left to give *wareni*.<sup>18</sup> Copying can then apply, because the application of (15) creates a derived environment; hence *wereni* results. Final Vowel Deletion gives [wɛrɛn], the correct surface form. In the case of /mware+ni/, (15) does not apply because the string fails to meet the structural description of the rule, while (16) is prohibited from applying by the RAC. That is, the sequence /aCɛ/ does not arise as a result of the combination of morphemes, nor does it arise as a result of the application of a previous rule. Therefore, \_\_Ce is a nonderived environment, and the application of the rule is blocked (hence Nonderived Environment Blocking).

Rule (16) was revised, as compared to its formulation in (4), to account for the data in (17), where *i* is transparent with respect to Copying. The forms in (17) also evidence the need for an additional rule—one of Glide Deletion—which may be expressed as in (18):

(17) Free Form	Gloss	Base Form	UR:Construct Form	SR:Construct Form
<i>pali</i>	'side'	<i>paliya</i>	<i>paliya+ni</i>	<i>pɛliyen</i>
<i>sali</i>	'meat'	<i>saliya</i>	<i>saliya+ni</i>	<i>sɛliyen</i>

(18) Glide Deletion

$$\left[ \begin{array}{c} \text{-syllabic} \\ \alpha F \end{array} \right] \rightarrow \emptyset / \left[ \begin{array}{c} \text{+syllabic} \\ \alpha F \end{array} \right] \_\_\_\#$$

Rule (18) applies after Final Vowel Deletion, deleting *y* after *i* and *w* after *u*, consistent with the observation that Pohnpeian exhibits no surface sequences of the shape /...iy#/ or /...uw#/. Glide deletion is also exhibited in (19):

(19) Base Form	<i>pati</i>	'eyebrow'
1SG	<i>pɛti</i>	
2SG	<i>pɛtim<sup>w</sup></i>	
3SG	<i>pati</i>	
1DU/PL EXC	<i>patit'</i>	
1DU INC	<i>patit'a</i>	
2DU	<i>patim<sup>w</sup>a</i>	
3DU	<i>patira</i>	
1PL INC	<i>patit'ayl</i>	
2PL	<i>patim<sup>w</sup>ayl</i>	
3PL	<i>patirayl</i>	
Construct	<i>pɛtin</i>	

Therefore, the first person singular form of this paradigm, *pɛti*, comes from underlying /pati+yil/, to which the rules of Low Vowel Raising, Final Vowel Deletion and Glide Deletion apply, in that order. This form poses no difficulty. However, the third person singular form *pati*, the underlying form of which has been hypothesised to be /pati+i/, cannot be accounted for within the analysis presented thus far; the first vowel of this form remains low, even though the underlying string meets the structural description of Low Vowel Raising and complies with the RAC. Note that /pati+yil/ yields *pɛti* and /pati+ni/ yields *pɛtin*, but /pati+i/ yields *pati*, not \**pɛti*. It might be possible to jury-rig a solution to this problem. For example, one might propose that a C must be present between *a* and *i*, but this seems hopelessly wrongheaded and totally unmotivated. Other possibilities suggest themselves, all equally implausible, so I will not entertain them here. What in fact seems to be the case is that either

<sup>18</sup> I assume that the directionality of this rule, as well as the rule of Vowel Copying, which applies iteratively, is predictable, as argued in Howard (1972).

the rule of Low Vowel Raising is wrong in some fundamental way, or the underlying representations to which it applies are faulty.

It is the latter possibility that, in fact, seems most likely. Consider that in Rehg 1986 (§4.2.3), it is argued that the only final short vowels that are retained by verbs with polysyllabic bases are /i/ and /ɛ/. The smart solution presented here, in part mirroring the historical facts of the language, hypothesises that, in nouns, short *a* also occurs in base-final position, thus suggesting asymmetrical phonological developments for these two major word classes. Further, the underlying representation /-i/ for the third person singular possessive pronoun is also difficult to justify. An underlying high vowel is proposed for this suffix, as well as for the first and second person singular and construct suffixes, primarily to motivate the raising and copying that occurs in association with them. Compare, however, the underlying representations of these forms with their Proto Micronesian (PMc) reconstructions.<sup>19</sup>

(20)	Base Forms	PMc
1SG	- <i>yi</i>	*- <i>xu</i>
2SG	- <i>m<sup>w</sup>i</i>	*- <i>m<sup>w</sup>u</i>
3SG	- <i>i</i>	*- <i>ñā</i>
Construct	- <i>ni</i>	*- <i>ni</i>

While none of these suffixes is ever followed by another suffix, thus creating an environment which preserves the final vowel, the base form *-m<sup>w</sup>i* can be justified on the grounds that (i) PMc \**u* is regularly reflected as Pohnpeian *i* in word-final position,<sup>20</sup> and (ii) this is the surface form of this suffix in honorific constructions (e.g. *m<sup>w</sup>arem<sup>w</sup>i* 'your (honorific) title'). The base form /-ni/ also appears defensible given the surface form of this suffix in compound words like *etiniyey* 'smoke' (from underlying /ati+ni+ayi/, literally 'current of fire'). While the final vowel of the first person singular form never occurs on the surface, the base form *-yi*, abstract though it might be, is at least feasible, since this base, like /-m<sup>w</sup>i/ and /-ni/ reflects regular sound correspondences and is consistent with its postulated protoform. But, within the smart solution, the only motivation for establishing /-i/ for the third person singular suffix is the fact that it triggers Raising and Copying. There is no other justification within Pohnpeian.<sup>21</sup> The final vowel of this suffix never surfaces, and this form is not cognate with PMc \*-ñā; /-i/ thus seems excessively abstract and cannot be justified on historical grounds.

In consideration of these substantive problems with the smart solution, a search for an alternative solution is obviously well motivated.

<sup>19</sup> The Proto Micronesian reconstructions are from Bender et al. (1984) and Jackson (1983, 1986).

<sup>20</sup> See Rehg (1973) for further discussion.

<sup>21</sup> Elbert (1974:35), however, observes that Puluwat has /-n/ in alternation with /-y/ for third person singular (and /-n/ ~ /-y/ ~ /-i/ for the construct). He states, "Third singular -n and -y are in free variation, with perhaps -y more common in connected discourse, but -n more so in citation forms". It is not yet clear how this observation relates to Pohnpeian.

#### 4 The dumb solution

As the term is employed here, and from the point of view of standard generative phonological theory, a **dumb** solution is one that ignores the requirement(s) that a single underlying representation be established for each morpheme and/or that the surface representations of such morphemes be derived through the application of well-motivated phonological rules. Such a dumb solution, might, in fact, take many forms, but only two are considered here.

The most extreme of these 'dumb' solutions is the **null solution**, in which one takes the position that all possessive paradigms are simply memorised by the native speaker, thus totally eliminating the need for rules of any sort. Although this solution might to some seem unworthy of consideration, it is at least worth mentioning, since the issue of determining what is memorised and what is rule-governed is in fact at the very heart of the linguistic enterprise. The problem with the null solution, of course, is that it is a solution of last resort. It is normally employed only when there is no regularity in the system, and, in the case of possessive paradigms, this is not the case. Regularity exists. It must exist, because it is extremely unlikely that any native speaker of Pohnpeian would, even in the course of a lifetime, hear all of the possessed forms of the hundreds of nouns that may be so inflected.<sup>22</sup> Yet, there is widespread agreement among native speakers of the language as to how such possessed forms should be pronounced. One can, of course, appeal to a notion like analogy to account for this fact, and postulate a limited number of exemplary paradigms that serve as models for other paradigms, but analogy implies regularity, and it is the task of characterising this regularity that is the central undertaking of modern linguistics. The approach that I have taken in this paper as a means of characterising such regularity is generative phonology. Others, of course, are possible.

A second 'dumb' solution is one that might be termed the **reanalysis solution**, which essentially entails making different morphological cuts than those made in the smart solution. In this solution, the final short vowels of noun bases are assumed to be the initial vowels of the possessive suffixes, presumably as a result of historical reanalysis by speakers of the language. The consequences of such an approach are (i) that possessive suffixes now have several allomorphs, and (ii) that nouns must be marked for membership in one of four paradigmatic classes. Examples of nouns belonging to each of these four classes, along with the allomorphs of the possessive suffixes one would need to postulate in such a solution, are shown in (21). Class 1 nouns are thus those that were previously analysed as being /i/ final, Class 2 nouns as /e/ final, Class 3 as /a/ final and Class 4 as long-vowel final.

(21) Base Forms	<i>pat</i>	<i>m<sup>w</sup>ar</i>	<i>mas</i>	<i>nt'aa</i>
	'eyebrow'	'title'	'face'	'blood'
Noun Classes	1	2	3	4
1SG	<i>-i</i>	<i>-ey</i>	<i>-ey</i>	<i>-y</i>
2SG	<i>-im<sup>w</sup></i>	<i>-em<sup>w</sup></i>	<i>-em<sup>w</sup></i>	<i>-m<sup>w</sup></i>
3SG	<i>-i</i>	<i>-e</i>	<i>-e</i>	$\emptyset$
1DU/PL EXC	<i>-it'</i>	<i>-et'</i>	<i>-at'</i>	<i>-t'</i>
1DU INC	<i>-it'a</i>	<i>-at'a</i>	<i>-at'a</i>	<i>-t'a</i>
2DU	<i>-im<sup>w</sup>a</i>	<i>-am<sup>w</sup>a</i>	<i>-am<sup>w</sup>a</i>	<i>-m<sup>w</sup>a</i>

<sup>22</sup> In my present dictionary database, consisting of more than 8,000 entries, there are 334 nouns that can be directly possessed.

3DU	-ira	-ara	-ara	-ra
1PL INC	-it'ayl	-at'ayl	-at'ayl	-t'ayl
2PL	-im <sup>w</sup> ayl	-am <sup>w</sup> ayl	-am <sup>w</sup> ayl	-m <sup>w</sup> ayl
3PL	-irayl	-arayl	-arayl	-rayl
Construct	-in	-en	-en	-n

While a reanalysis solution similar to this one enabled Hale (1973) to insightfully account for the shape of transitive suffixes in Maori, such an approach accomplishes little for Pohnpeian. Although it is true that the base forms listed above are less abstract than those set up within the smart solution, it is still not clear how one is to formulate the Raising and/or Copying rules required to explain alternations in vowel quality in nominal bases. In fact, the problem is greatly exacerbated. For example, note in Class 1 that the first person singular and third person singular forms are identical—both are *-i*. Yet, the first person singular form triggers raising (*peti* ‘my eyebrow’), while the third person singular form does not (*piti* ‘his/her/its eyebrow’). Also note that the singular and construct forms in Classes 2 and 3 are identical, but nouns of these two classes do not behave identically in combination with these suffixes. Compare *m<sup>w</sup>arey* ‘my title’ with *mese<sup>y</sup>* ‘my face’. The rule(s) that would need to be formulated to account for the alternations between /a/ and /e/ within this solution would have to rely extensively on morphological information, thus violating at least the spirit of generative phonology. Still, as will be noted in the next section, the reanalysis solution is not totally without merit.

## 5 The Pohnpeian solution

When native speakers of Pohnpeian are asked to divide directly possessed nouns into their meaningful parts, their responses are nearly always the same.<sup>23</sup> They express considerable uncertainty over where to make the cuts in paradigms of the types represented by Classes 1, 2 and 4 in (21), but in my experience, they universally make the same cuts as in the paradigm represented by Class 3. This observation thus suggests a third possible solution, called here the **Pohnpeian solution**.

In this solution, it is assumed that all possessive suffixes have two allomorphs, as listed below. The first set of allomorphs occurs after vowel-final bases and the second after consonant-final bases. The number markers that occur with the nonsingular forms are *-aa* ‘dual’ and *-ayl* ‘plural’.

(22)	After Vowels	After Consonants
1SG	-yi	-ey
2SG	-m <sup>w</sup> i	-em <sup>w</sup>
3SG	-∅	-εε <sup>24</sup>
1nonSG	-t'	-at'
2nonSG	-m <sup>w</sup>	-am <sup>w</sup>

<sup>23</sup> Although I have not conducted a formal study of how Pohnpeians believe such forms should be divided, I have posed this question to probably no fewer than 30 native speakers, with remarkably consistent results.

<sup>24</sup> Final Vowel Deletion motivates positing a long vowel /-εε/ in the third person singular form for the suffix set employed after consonants.



3nonSG	-r	-ar
Construct	-ni	-en

As in the 'smart' solution, the two rules given earlier as (15) and (16) are required to account for the raising of /a/ to /ε/; (15) and (16) are repeated here as (23) and (24):

(23) Low Vowel Raising  
 $\left[ \begin{array}{c} a \\ \text{-long} \end{array} \right] \rightarrow \epsilon / \_ (C_0i)C_0i$

(24) Vowel Copying  
 $\left[ \begin{array}{c} a \\ \text{-long} \end{array} \right] \rightarrow \epsilon / \_ (C_0i)C_0\epsilon$

As previously noted, it would appear that these two rules should be collapsed into one. This is apparently not possible, however, since these rules in fact have different domains of application.<sup>25</sup> In my dissertation, I encoded domain differences in terms of boundaries. In my current research, conducted within the framework of lexical phonology, I argue that these rules differ in terms of the strata in which they apply. That is, Low Vowel Raising applies only in Stratum 1, where possessive suffixes are attached, whereas Vowel Copying applies in both Strata 1 and 2, the latter being where numeral classifiers are combined with numeral stems. Note, therefore, that Low Vowel Raising does **not** occur in the numerals.<sup>26</sup> Vowel Copying, however, does apply in such forms, as (26) attests:

(25) UR	SR	Gloss
<i>ria+kisi</i>	<i>riakis</i>	'two (small pieces)'
<i>lima+kisi</i>	<i>limakis</i>	'five (small pieces)'
<i>tuwa+kisi</i>	<i>tuwakis</i>	'nine (small pieces)'
<i>ria+tipe</i>	<i>riatip</i>	'two (slices)'
<i>lima+tipe</i>	<i>limatip</i>	'five (slices)'
<i>tuwa+tipe</i>	<i>tuwatip</i>	'nine (slices)'
(26) <i>ria+meni</i>	<i>riemen</i>	'two (animate beings)'
<i>ria+lepi</i>	<i>rielep</i>	'two (oblong pieces)'
<i>ria+t'εε</i>	<i>riet'ε</i>	'two (leaves)'

The sample derivations in (27) illustrate how these rules apply in association with the forms of the possessive pronouns listed in (22).

(27) Gloss	'my eyebrow'	'my title'	'my face'	'my blood'	'side of'
UR	<i>pai+yi</i>	<i>m<sup>w</sup>are+yi</i>	<i>mas+εy</i>	<i>ni'aa+yi</i>	<i>paliy+en</i>
Raising	<i>peti+yi</i>	--	--	--	--
Copying	--	--	<i>mes+εy</i>	--	<i>peliy+en</i>
Final V Deletion	<i>peti+y</i>	<i>m<sup>w</sup>are+y</i>	--	<i>ni'aa+y</i>	--
Glide Deletion	<i>peti</i>	--	--	--	--
SR	<i>peti</i>	<i>m<sup>w</sup>arey</i>	<i>mesey</i>	<i>ni'a:y</i>	<i>peliyen</i>

In this analysis, the third person singular allomorph that occurs after bases ending in vowels is represented by -∅. A zero allomorph is necessary to block Final Vowel Deletion in forms like

<sup>25</sup> See §6.3 of Rehg 1986 for further discussion.

<sup>26</sup> For the purposes of this paper, I will continue to use + to mark morpheme boundary, rather than the bracket notation of lexical phonology. The distinction has no bearing on the argument being developed here.



those in (28). The need to posit a zero allomorph is obviously a contentious part of this analysis, but the substantive issues that arise in association with positing such an element will not be further pursued here.<sup>27</sup>

(28) UR	<i>pai+∅</i>	<i>m<sup>w</sup>are+∅</i>	<i>nt'aa+∅</i>
Final V Deletion	--	--	--
SR	<i>pai</i>	<i>m<sup>w</sup>are</i>	<i>nt'a:</i>

There are a number of reasons to believe that the **Pohnpeian solution** is at least on the right track. First, the morpheme cuts are consistent with native-speaker intuitions. The facts of the language apparently giving rise to such intuitions are illustrated by the paradigms in (29) and (30), the first for the consonant-final base /mas/ 'face' and the second for the general possessive classifier /a-/.<sup>28</sup>

(29)	'face'	'general possessive classifier'
1SG	<i>mesey</i>	<i>ey (a:y)</i>
2SG	<i>mesem<sup>w</sup></i>	<i>em<sup>w</sup> (a:m<sup>w</sup>)</i>
3SG	<i>mesε</i>	<i>ε (a: ~ ε:)</i>
1DU/PL EXC	<i>masat'</i>	<i>at' (a:t')</i>
1DU INC	<i>masat'a</i>	<i>at'a</i>
2DU	<i>masam<sup>w</sup>a</i>	<i>am<sup>w</sup>a</i>
3DU	<i>masara</i>	<i>ara</i>
1PL INC	<i>masat'ayl</i>	<i>at'ayl</i>
2PL	<i>masam<sup>w</sup>ayl</i>	<i>am<sup>w</sup>ayl</i>
3PL	<i>masarayl</i>	<i>arayl</i>
Construct	<i>mesεn</i>	<i>εn</i>

The possessive suffixes that follow a consonant-final base like /mas/ are virtually identical to the forms of the general possessive classifier and, indeed, native speakers of the language consciously make this association. Note also that the monosyllabic forms of the general possessive classifier (the first four forms above) exhibit both short and long forms (the latter in parentheses). The long forms are those that occur alone in a noun phrase (e.g. *ey p<sup>w</sup>owd* 'my spouse' vs *a:y* 'mine'). The fact that three forms exist for the third person singular form of the general possessive classifier is, I believe, a consequence of the fact that speakers associate this form with the third person singular form of the possessive suffix employed after consonant-final noun bases, but explaining the relationships among these forms is a complex task that is beyond the scope of this paper.

Second, crucial evidence exists to support the claim that not only **might** the Pohnpeian solution be correct, but that it **must** be correct. Consider, therefore, the paradigms in (30):

(30) Base Forms	<i>m<sup>w</sup>are</i>	'title'	<i>saleŋ</i>	'ear'
1SG	<i>m<sup>w</sup>arey</i>		<i>saleŋey</i>	
2SG	<i>m<sup>w</sup>arem<sup>w</sup></i>		<i>saleŋem<sup>w</sup></i>	
3SG	<i>m<sup>w</sup>are</i>		<i>saleŋe</i>	

<sup>27</sup> One might also posit a timing slot that is unassociated with a melody. This strategy, in fact, seems justified based upon my claims about the foot structure of the language, a topic I intend to explore in a subsequent paper.

<sup>28</sup> Among all the classifiers, the general possessive classifier almost certainly occurs with the greatest frequency, though I have no statistics to support this claim.

1 DU/PL EXC	<i>m<sup>w</sup>arɛt'</i>	<i>salanɔt'</i>
1 DU INC	<i>m<sup>w</sup>arat'a</i>	<i>salanɔt'a</i>
2 DU	<i>m<sup>w</sup>aram<sup>w</sup>a</i>	<i>salanam<sup>w</sup>a</i>
3 DU	<i>m<sup>w</sup>arara</i>	<i>salanara</i>
1 PL INC	<i>m<sup>w</sup>arat'ayl</i>	<i>salanɔt'ayl</i>
2 PL	<i>m<sup>w</sup>aram<sup>w</sup>ayl</i>	<i>salanam<sup>w</sup>ayl</i>
3 PL	<i>m<sup>w</sup>ararayl</i>	<i>salanarayl</i>
Construct	<i>m<sup>w</sup>aren</i>	<i>salenɛn</i>

An underlying /*ɛ*/ is postulated in both base forms above since, as noted previously, the presence of this vowel blocks copying. Assuming that these base forms are correct, a phenomenon alluded to in §3 is the lowering of /*ɛ*/ to /*a*/ in the nonsingular forms of these paradigms.

The issue here, then, is how the rule of /*ɛ*/ Lowering is to be stated. The paradigms above suggest that it might be formulated so that /*ɛ*/ is lowered to /*a*/ when flanked by syllables containing /*a*/; therefore, it might be stated as  $\epsilon \rightarrow a / aC\_Ca$ . But the paradigms in (31) illustrate that this is not the case:

(31) Base Forms	<i>ɲile</i>	'voice'	<i>nsɛnɛ</i>	'will'
1 SG	<i>ɲiley</i>		<i>nsɛney</i>	
2 SG	<i>ɲilem<sup>w</sup></i>		<i>nsɛnem<sup>w</sup></i>	
3 SG	<i>ɲile</i>		<i>nsɛnɛ</i>	
1 DU/PL EXC	<i>ɲilet'</i>		<i>nsɛnet'</i>	
1 DU INC	<i>ɲilat'a</i>		<i>nsɛnata'</i>	
2 DU	<i>ɲilam<sup>w</sup>a</i>		<i>nsɛnam<sup>w</sup>a</i>	
3 DU	<i>ɲilara</i>		<i>nsɛnara</i>	
1 PL INC	<i>ɲilat'ayl</i>		<i>nsɛnat'ayl</i>	
2 PL	<i>ɲilam<sup>w</sup>ayl</i>		<i>nsɛnam<sup>w</sup>ayl</i>	
3 PL	<i>ɲilarayl</i>		<i>nsɛnarayl</i>	
Construct	<i>ɲilen</i>		<i>nsɛnɛn</i>	

Based on these data, a preferable statement of the rule of /*ɛ*/ Lowering is provided in (32). This rule states that /*ɛ*/ lowers to *a* in assimilation to a following /*a*/. Since this rule is an obligatory neutralisation rule, it is constrained by the RAC. Therefore, Pohnpeian exhibits surface occurrences of /*ɛ*/ followed by /*a*/, as illustrated in (33). While it is clear that some of these forms were morphologically complex earlier in the history of the language, no reason exists to treat them as such synchronically.

(32) /*ɛ*/ Lowering  
 $\epsilon \rightarrow a / \_ + C_0a$

(33) <i>metaka:n</i>	'talkative'
<i>kemmat</i>	'put on dry clothing'
<i>kempak</i>	'shallow'
<i>tekatek</i>	'thorny'
<i>lepalep</i>	'to doze'

Unlike the rule of Vowel Copying, /*ɛ*/ Lowering does not apply iteratively (as illustrated by forms like *nsɛnat'a*); hence, in the informal notation employed in (32), the rule is anchored

by the presence of a morpheme boundary. Like the rule of Vowel Copying, however, this rule applies, not only in association with possessive suffixes, but with numeral classifiers as well:<sup>29</sup>

(34) UR	SR	Gloss
<i>wenɛ+paa</i>	<i>wenapa</i>	'six (fronds)'
<i>wenɛ+kap</i>	<i>wenakap</i>	'six (bundles)'
<i>wenɛ+kaa</i>	<i>wenaka</i>	'six (rows)'

Sample derivations illustrating how this rule interacts with first person dual/plural exclusive and dual inclusive suffixes of both allomorph classes are provided in (35).<sup>30</sup>

(35) UR	<i>m<sup>w</sup>are+t'</i>	<i>m<sup>w</sup>are+t'+aa</i>	<i>saleŋ+at'</i>	<i>saleŋ+at'+aa</i>
/ɛ/ Lowering	--	<i>m<sup>w</sup>ara+t'+aa</i>	<i>saleŋ+at'</i>	<i>saleŋ+at'+aa</i>
Final V Deletion	--	<i>m<sup>w</sup>ara+t'+a</i>	<i>saleŋ+at'</i>	<i>saleŋ+at'+a</i>
SR	<i>m<sup>w</sup>aret'</i>	<i>m<sup>w</sup>arat'a</i>	<i>saleŋat'</i>	<i>saleŋat'a</i>

The forms in (35) illustrate that, if morpheme cuts were made in a manner consistent with the smart solution, in which all bases end in underlying vowels, it would not be possible to generate the correct surface forms for the paradigm for 'ear', as illustrated in (36). Since /ɛ/ Lowering cannot apply morpheme-internally (it is constrained by the RAC), the vowel that triggers this lowering **must** be in the next morpheme. The Pohnpeian solution is thus strongly supported by these data.

(36) UR	<i>saleŋa+t'</i>	<i>saleŋa+t'+aa</i>
/ɛ/ Lowering	--	--
Final V Deletion	--	<i>saleŋa+t'+a</i>
SR	<i>*saleŋat'</i>	<i>*saleŋat'a</i>

Two additional paradigms not yet considered provide still further support, albeit of a different nature, for the Pohnpeian solution. Specifically, they provide important sources of data that help us understand how borrowed words are treated in relation to the Pohnpeian solution. Consider, therefore, the nouns in (37), which clearly derive from foreign sources. The Pohnpeian word for 'movie' is from Japanese *katsudoo*, while the word for 'book', *p<sup>w</sup>u:k*, is from English.<sup>31</sup> The possessed forms of these nouns mean 'movie about me', 'book about me', 'movie about you', 'book about you' etc.

(37) Free Form	<i>kasto</i>	'movie'	<i>p<sup>w</sup>u:k</i>	'book'
Base Form	<i>kastoo</i>		<i>p<sup>w</sup>uk</i>	
1SG	<i>kasto:y</i>		<i>p<sup>w</sup>ukɛy</i>	
2SG	<i>kasto:m<sup>w</sup></i>		<i>p<sup>w</sup>ukɛm<sup>w</sup></i>	
3SG	<i>kasto:</i>		<i>p<sup>w</sup>uke</i>	
1DU/PL EXC	<i>kasto:t'</i>		<i>p<sup>w</sup>ukat'</i>	
1DU INC	<i>kasto:t'a</i>		<i>p<sup>w</sup>ukat'a</i>	
2DU	<i>kasto:m<sup>w</sup>a</i>		<i>p<sup>w</sup>ukam<sup>w</sup>a</i>	

<sup>29</sup> This rule also applies in association with verbal suffixes, but those data will not be considered here.

<sup>30</sup> No crucial evidence for the ordering of /ɛ/ Lowering and Final Vowel Deletion are exhibited in these data; however, based upon other facts not considered here, it seems clear that Final Vowel Deletion applies in the last lexical stratum.

<sup>31</sup> Borrowed words that can be directly possessed are uncommon. This is to be expected, however, given the 'inalienable' character of such nouns.

3DU	<i>kasto:ra</i>	<i>p<sup>w</sup>ukara</i>
1PL INC	<i>kasto:t'ayl</i>	<i>p<sup>w</sup>ukat'ayl</i>
2PL	<i>kasto:mwayl</i>	<i>p<sup>w</sup>ukam<sup>w</sup>ayl</i>
3PL	<i>kasto:rayl</i>	<i>p<sup>w</sup>ukarayl</i>
Construct	<i>kasto:n</i>	<i>p<sup>w</sup>uken</i>

The first paradigm, that for *kasto*, sheds light on how Pohnpeian speakers establish underlying representations for loan words that end in vowels. It seems clear that such words are interpreted as having underlying final long vowels, as indeed the word *kasto* does in Japanese, the source language. However, virtually all loan words that have final long vowels in Japanese have final short vowels in Pohnpeian. That is, they are subject to the Pohnpeian rule of Final Vowel Deletion, which has the effect of shortening long vowels in word-final position.<sup>32</sup> When followed by an affix, however, as in (37), such vowels remain long.

More interesting is the word for 'book'. Within the framework of the smart solution, it would be necessary to establish /pwuka/ as the base form of this noun. But this word is obviously a borrowing from English. A final vowel in this base cannot be justified on the basis of either the synchronic form of this word in English or the diachronic form of this word in Pohnpeian. Further, within the smart solution, the fact that this base ends in /a/ as opposed to /i/ or /e/ is inexplicable. The Pohnpeian solution, however, encounters no such problems. Within this analysis, it is assumed that the base is consonant-final, as it is in the donor language, and, as a consequence, the possessive paradigm for 'book' **must** be as it is. Consequently, one of the desiderata of generative phonology—that there be a single underlying representation for each morpheme—must be compromised to account for the complex patterns of vowel alternations in Pohnpeian.

## References

- Abaglo, P. and Diana Archangeli, 1989, Language particular underspecification: Gengbe /e/ and Yoruba /i/. *Linguistic Inquiry* 20:457–480.
- Bender, Byron W., 1973, Parallelisms in the morphophonemics of several Micronesian languages. *Oceanic Linguistics* 1/1&2:455–477.
- Bender, Byron W., Robert W. Hsu, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel and Judith W. Wang, 1984, Micronesian cognate sets. Computer printout. Honolulu, Department of Linguistics, University of Hawai'i.
- Chomsky, Noam and Morris Halle, 1968, *The sound pattern of English*. Cambridge, MA: MIT Press.
- Cole, Jennifer, 1995, The cycle in phonology. In John A. Goldsmith, ed. *The handbook of phonological theory*, 70–113. Cambridge MA: Blackwell.
- Elbert, Samuel H., 1974, *Puluwat grammar*. Canberra: Pacific Linguistics.
- Goldsmith, John A., ed., 1999, *Phonological theory: the essential readings*. Malden, Mass.: Blackwell.

<sup>32</sup> See Miyagi 2000 for further discussion. It should be noted that Pohnpeian has no prohibition against surface final long vowels.

- Hale, Kenneth, 1973, Deep-surface canonical disparities in relation to analysis and change: an Australian example. In Thomas A. Sebeok, ed. *Current trends in linguistics*, vol. 11: *Diachronic, areal, and typological linguistics*, 401–458. The Hague: Mouton.
- Howard, Irwin Jay, 1972, A directional theory of rule application in phonology. PhD dissertation, Massachusetts Institute of Technology.
- Jackson, Frederich H., 1983, The internal and external relationships of the Trukic languages of Micronesia. PhD dissertation, University of Hawai‘i.
- 1986, On determining the external relationships of the Micronesian languages. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, 201–238. Canberra: Pacific Linguistics.
- Iverson, Gregory and Deirdre Wheeler, 1988, Blocking and the elsewhere condition. In M. Hammond and M. Noonan, eds *Theoretical morphology: approaches in modern linguistics*, 325–338. San Diego: Academic Press.
- Kiparsky, Paul, 1968, *How abstract is phonology?* Indiana University Linguistics Club. Reprinted in 1973 as Abstractness, opacity, and global rules. In Osamu Fujimura, ed. *Three dimensions of linguistic theory*. Tokyo: TEC.
- 1982, From cyclic phonology to lexical phonology. In H. van der Hulst and N. Smith, eds *The structure of phonological representations*, vol. 1, 131–175. Dordrecht: Foris. Portions reprinted in 1999 in John Goldsmith, ed. *Phonological theory: the essential readings*, 34–62.
- 1993, Blocking in non-derived environments. In S. Hargus and E. Kaisse, eds *Studies in lexical phonology*. San Diego: Academic Press.
- Mascaró, Joan, 1976, Catalan phonology and the phonological cycle. PhD dissertation, Massachusetts Institute of Technology.
- Miyagi, Kimi, 2000, Japanese loanwords in Pohnpeian: adaptation and attrition. *Japanese Linguistics* 7:114–131. Tokyo: The National Language Research Institute.
- Rehg, Kenneth L., 1973, On the history of Ponapean phonology. *Working Papers in Linguistics* 5(8):17–55. Honolulu: Department of Linguistics, University of Hawai‘i.
- 1984, The origins of ‘compensatory lengthening’ rules in Micronesian languages. In Byron W. Bender, ed. *Studies in Micronesian linguistics*, 53–59. Canberra: Pacific Linguistics.
- 1986, The phonology of Ponapean vowels. PhD dissertation, University of Hawai‘i.
- 1993, Proto-Micronesian prosody. In Jerold A. Edmondson and Kenneth J. Gregerson, eds *Tonality in Austronesian languages*, 25–46. Oceanic Linguistics Special Publication No. 24. Honolulu: University of Hawai‘i Press.
- Rehg, Kenneth L. and Damian G. Sohl, 1979, *Ponapean–English dictionary*. Honolulu: University of Hawai‘i Press.
- 1981, *Ponapean reference grammar*. Honolulu: University of Hawai‘i Press.
- Steriade, Donca, 1995, Underspecification and markedness. In John A. Goldsmith, ed. *The handbook of phonological theory*, 114–174. Cambridge, Mass.: Blackwell.



# *On the development of agreement markers in some Northern Philippine languages*

---

LAWRENCE A. REID

This paper attempts to provide an explanation for an innovation occurring in the Central Cordilleran languages of the Philippines, in which what were originally derivational endings *en* and *an* lose their final nasal when they carry first, second or third person singular agreement features, respectively *k*, *m* and *na*. It is claimed that this situation arose as a result of the incorporation of a reduced form of what was originally a genitive case-marking preposition *ni* as *n* onto vowel-final verbs and their nominal counterparts. An analogy was then established between these forms ending in *n* and derived forms with *en* and *an* endings, so that the latter were also perceived as being vowel-final for the purpose of substituting the *k*, *m* and *na* agreement endings. The discussion is presented within the Lexicase theoretical framework, specifically its claim that words have neither internal structure nor morphological boundaries.

## 1 Introduction<sup>1</sup>

Alternation in the forms of the first and second person singular, so-called clitic genitive pronouns in a number of Austronesian languages has been noted in various publications, including Dyen (1974), Tharp (1974), Blust (1977) and Reid (1979, 1981). The alternation is found in all but a few of the Cordilleran languages of the Northern Philippines,<sup>2</sup> where the full

---

<sup>1</sup> It is a privilege to be invited to contribute a paper to this volume honouring Byron, who for decades has introduced students to the complexities of Philippine verbal morphology through a series of exercises on Tagalog that formed part of his Introduction to Morphology course, and who through his own quiet style has encouraged me in my own explorations of Philippine languages. I would like to thank Ritsuko Kikusawa, Hsiu-chuan Liao, Carl Rubino and Stan Starosta for reading and commenting on this paper. This analysis is presented within the framework and notational conventions of Lexicase (Starosta 1988 and subsequent works) and seamless morphology (Ford, Singh & Martohardjono 1997 and others discussed in §3). All errors of fact, interpretation and application of the theory, however, are unfortunately mine alone.

<sup>2</sup> The alternation is found in all of the Central and Southern Cordilleran languages, in both Northern and Southern Alta (Reid 1991), in Ilokano (Rubino 1997), and in the Cagayan Valley languages, but not in the

forms are typically =*ku* '1s' and =*mu* '2s'<sup>3</sup> following consonant-final stems. Vowel-final stems, however, take reduced =*k* '1s' and =*m* '2s', respectively. These alternations recur, regardless of whether the form is functioning as the possessor of a noun or as the agent of a transitive verb, as shown in Tables 1 and 2, which demonstrate the complete set of genitive pronominal forms in Bontok, a Central Cordilleran language.<sup>4</sup> The forms are listed in these tables following the usual analysis of such forms in the literature, but without at this point making any commitment to their actual status.<sup>5</sup>

**Table 1:** Bontok Genitive possessive forms

	POSTCONSONANTAL		POSTVOCALIC	
1S	<i>ʔábuŋ =kú</i>	'my house'	<i>ʔásu =k</i>	'my dog'
2S	<i>ʔábuŋ =mú</i>	'your house'	<i>ʔásu =m</i>	'your dog'
3S	<i>ʔábuŋ =ná</i>	'his/her house'	<i>ʔásu =ná</i>	'his/her dog'
1+2S	<i>ʔábuŋ =tá</i>	'our (DL) house'	<i>ʔásu =tá</i>	'our (DL) dog'
1P	<i>ʔábuŋ =mí</i>	'our (EX) house'	<i>ʔásu =mí</i>	'our (EX) dog'
1+2P	<i>ʔábuŋ =takú</i>	'our (IN) house'	<i>ʔásu =takú</i>	'our (IN) dog'
2P	<i>ʔábuŋ =yú</i>	'your (PL) house'	<i>ʔásu =yú</i>	'your (PL) dog'
3P	<i>ʔábuŋ =dá</i>	'their house'	<i>ʔásu =dá</i>	'their dog'

Negrito languages of Northeastern Luzon (Headland & Headland 1974, Reid 1983), nor in Arta, a Negrito language isolate spoken in the Cagayan Valley (Reid 1989). Yogad, one of the Northern Cordilleran languages, has a variant following vowel-final words only for the second singular genitive pronoun (Healey 1958).

<sup>3</sup> An equals (=) sign preceding a form marks it as an enclitic.

<sup>4</sup> All Bontok forms are cited in phonemic transcription from the author's fieldnotes.

<sup>5</sup> Two sets of abbreviations are used in this paper. The set used in conventional Lexicase feature matrices includes: *addr* – addressee, *AGT* – Agent, *cltc* – clitic, *COR* – Correspondent, *F* – semantic feature, *Gen* – Genitive, *N* – Noun, *plrl* – plural, *prnn* – pronoun, *pssd* – possessed, *spkr* – speaker, *trns* – transitive. Other abbreviations used in glosses and elsewhere include: *1S* – first person singular pronoun, *1P* – first person exclusive pronoun, *1+2S* – first person dual pronoun, *1+2P* – first person plural inclusive pronoun, *2S*, second person singular pronoun, *2P* – second person plural pronoun, *3S* – third person singular pronoun, *3P* – third person plural pronoun, *Det* – Determiner, *DL* – dual, *EX* – exclusive, *IN* – inclusive, *NP* – noun phrase, *P* – Preposition, *PL* – plural, *PPh* – Proto Philippine, *SG* – singular, *s.th.* – something, *WFS* – word formation strategy.



**Table 2:** Bontok Genitive agent forms

	POSTCONSONANTAL		POSTVOCALIC	
1S	<i>ʃwasít =kú</i>	'I throw s.th. out'	<i>ʃláku =k</i>	'I sell s.th.'
2S	<i>ʃwasít =mú</i>	'you throw s.th. out'	<i>ʃláku =m</i>	'you sell s.th.'
3S	<i>ʃwasít =ná</i>	'he/she throws s.th. out'	<i>ʃláku =ná</i>	'he/she sells s.th.'
1+2S	<i>ʃwasít =tá</i>	'we (DL) throw s.th. out'	<i>ʃláku =tá</i>	'we (DL) sell s.th.'
1P	<i>ʃwasít =mí</i>	'we (EX) throw s.th. out'	<i>ʃláku =mí</i>	'we (EX) sell s.th.'
1+2P	<i>ʃwasít =takú</i>	'we (IN) throw s.th. out'	<i>ʃláku =takú</i>	'we (IN) sell s.th.'
2P	<i>ʃwasít =yú</i>	'you (PL) throw s.th. out'	<i>ʃláku =yú</i>	'you (PL) sell s.th.'
3P	<i>ʃwasít =dá</i>	'they throw s.th. out'	<i>ʃláku =dá</i>	'they sell s.th.'

In the Central Cordilleran languages,<sup>6</sup> as well as in Ilokano, an innovation has produced an additional environment in which the shortened forms are found. On transitive verbs containing a reflex of either \*-en or \*-an,<sup>7</sup> the short pronominal form replaces the final *-n* of the verb ending. In each of these languages (but not in Ilokano) the final *-n* is also replaced when the third person singular form =*ná* occurs, as shown in Table 3.

**Table 3:** Bontok Genitive Agent forms following transitive verbs with \*-en or \*-an suffixes

1S	<i>ʔaláʔə =k</i>	'I get s.th.'	<i>dalusá =k</i>	'I clean s.th.'
2S	<i>ʔaláʔə =m</i>	'you get s.th.'	<i>dalusá =m</i>	'you clean s.th.'
3S	<i>ʔaláʔə =ná</i>	'he/she gets s.th.'	<i>dalusá =ná</i>	'he/she cleans s.th.'
1+2S	<i>ʔaláʔən =tá</i>	'we (DL) get s.th.'	<i>dalusán =tá</i>	'we (DL) clean s.th.'
1P	<i>ʔaláʔən =mí</i>	'we (EX) get s.th.'	<i>dalusán =mí</i>	'we (EX) clean s.th.'
1+2P	<i>ʔaláʔən =takú</i>	'we (IN) get s.th.'	<i>dalusán =takú</i>	'we (IN) clean s.th.'
2P	<i>ʔaláʔən =yú</i>	'you (PL) get s.th.'	<i>dalusán =yú</i>	'you (PL) clean s.th.'
3P	<i>ʔaláʔən =dá</i>	'they get s.th.'	<i>dalusán =dá</i>	'they clean s.th.'

The primary purpose of this paper is to examine the factors that have brought about the use of the postvocalic variants on transitive verbs that otherwise would end in a consonant. However as a prelude to this discussion, it will be necessary to consider the status of the forms themselves.

## 2 Full words, clitics, affixes or none of the above?

Determining where words begin and end is of primary importance when one's theory of language claims that it is full words that are the units that are stored in one's lexicon and that each word carries with it all the information that is necessary not only for its semantic

<sup>6</sup> The Central Cordilleran languages include Bontok, Kankanaey, Balangaw and Ifugaw (the Nuclear Central Cordilleran languages), Kalinga and Itneg (which together with the Nuclear group constitute North Central Cordilleran) and Isinai (Reid 1974).

<sup>7</sup> My use of reconstructed affixes is inconsistent with the theoretical claims being made in this paper, namely that words have no internal structure. They are cited here as affixes solely for their heuristic value.

interpretation but also for its syntactic distribution. The problem here lies in determining whether a phonological word constitutes a single lexical item, or whether it is constituted of more than one lexical item, one of which is a free form, and the other a clitic having its own syntactic privileges but phonologically attached to the free form.

The problem of determining whether a phonologically bound form in a language is a clitic or an affix was the topic of work by Zwicky and Pullum (1983). A later article by Zwicky (1985) tackled the related problem of determining whether a given form is a clitic or an independent word. Each paper suggests a series of tests by which evidence for one or another analysis could be adduced. In this section, I will attempt to apply some of the tests proposed in those papers to determine whether the genitive pronominal forms of the kind shown in Tables 1–3 are actually clitics (as they are characterised in many descriptions of Philippine languages); whether they are independent lexical items (as they are usually written in local orthographies); or whether, at least in some cases, they are neither clitics nor separate words, but are forms that have been incorporated into their former host, in the process contributing some of their features to those of their host as agreement features.

## 2.1 Full words or clitics?

By several of Zwicky's tests most of the forms cited in the tables above are at least clitics and not free forms. One of the tests for clitics is that they form a phonological unit with an independent word (Zwicky 1985:286). The postvocalic forms *-k* and *-m* are by this criterion clearly not independent words. If they were, they would be the only words in the language with no vocalic nucleus of their own. It is not clear, however, that the longer forms, those having at least one vocalic nucleus, constitute a phonological unit with the independent word they follow, for the following three reasons:

- (i) Pronominal forms carry their own stress, even if they are only single syllables. Note that in Bontok, the sequence *ʔásu=ná<sub>1</sub>* 'his dog' is homophonous with *ʔásu ná<sub>2</sub>* 'this is a dog', in which *ná<sub>2</sub>* is a demonstrative noun freely substituting for a full NP such as *nan dáʔəy* 'this (one)', in a sentence such as *ʔásu nan dáʔəy* 'This one is a dog'.
  - (ii) Monosyllabic prepositions and determiners, on the other hand, are typically unstressed, so that the pronominal form *dá<sub>1</sub>* in the sequence *ʔásu=dá<sub>1</sub> nan lallaláki* 'the men's dog' (lit. 'their dog, the men') is not homophonous with the unstressed plural determiner *da<sub>2</sub>* in *ʔásu-n da<sub>2</sub>=Juán* 'the dog of John and his companions', where *da<sub>2</sub>* is phonologically (and syntactically) dependent on the word that follows it.
  - (iii) The presence of a pronominal form does not affect the position of stress on the preceding word. Words are stressed on either their ultimate or their penultimate syllable, so that a transitive verb that is suffixed carries stress one syllable further to the right than its associated form without a suffix, as shown in (1–4). Note that the verb meaning 'get' in (1) and (2) is always stressed on the penultimate syllable, while the verb meaning 'eat' in (3) and (4) is always stressed on the ultimate syllable, regardless of the presence or absence of a pronominal form.
- (1) a. *manála* 'to get ([-trns])'  
 b. *manála =ká* 'you (SG) get (some)'  
 c. *manála =kayú* 'you (PL) get (some)'

- (2) a. *ʔaláʔən* 'get ([+trns])'  
 b. *ʔaláʔən =yú* 'you (PL) get (it)'  
 c. *ʔaláʔən =takú* 'we (IN) get (it)'
- (3) a. *maŋán* 'to eat ([-trns])'  
 b. *maŋán =ká* 'you (SG) eat'  
 c. *maŋán =kayú* 'you (PL) eat'
- (4) a. *kanən* 'eat ([+trns])'  
 b. *kanən =yú* 'you (PL) eat (it)'  
 c. *kanən =takú* 'we (IN) eat (it)'

Zwicky further notes that "if an element counts as belonging to a phonological word for the purposes of accent, tone, or length assignment, then it should be a clitic" (1985:286). By this criterion, however, the pronominal forms should be clitics, not independent words, because in Bontok, as in many Philippine languages, only stressed open penultimate syllables have lengthened vowels. Open ultimate stressed syllables do not have long vowels. However, a form with a stressed open ultimate syllable does carry vowel length, if a pronominal form follows it, as in (5).

- (5) a. *matá* [matá] 'eye'  
 b. *maták* [maták] 'my eye'  
 c. *matá=ná* [matá:ná] 'his eyes'  
 d. *matá=yú* [matá:yú] 'your eyes'

A further test is stated as follows, "an element affected by or conditioning a sandhi rule otherwise known to be internal should be a clitic" (Zwicky 1985:286). This particular test is of little relevance to the Bontok data cited above, because there are no sandhi rules operating in the data. However if we look at data from Karao, a Southern Cordilleran language, we find clear evidence that postvocalic variants of genitive pronouns are, at least in this language, part of the preceding phonological word and are probably therefore clitics. Apart from the first and second person singular forms, which retain the Proto Cordilleran postvocalic variants =*k* and =*m* respectively, Karao uses a set of innovated intervocalic consonant-initial variants for each of the other forms.

**Table 4:** Karao Genitive<sup>8</sup> pronouns (Brainerd 1997:146)

	POSTCONSONANTAL	POSTVOCALIC
1S	<i>ko</i>	<i>k</i>
2S	<i>mo</i>	<i>m</i>
3S	<i>to</i>	<i>tho [θo]</i>
1+2S	<i>mi</i>	<i>wi</i>
1P	<i>tayo</i>	<i>thayo [θayo]</i>
1+2P	<i>tayocha</i>	<i>thayocha</i>
2P	<i>jo</i>	<i>yo</i>
3P	<i>cha</i>	<i>ra</i>

Finally, Zwicky notes, “We expect that bound elements will be affixes, but that free elements will constitute independent words. Correspondingly, if an element is bound, and especially if it cannot occur in complete isolation, it should be a clitic; if free, and especially if it occurs in complete isolation, it should be an independent word” (1985:287). By these criteria, none of the pronominal forms that we have been discussing can be considered to be free forms. None of them can occur independently from the form to which they are phonologically attached, and no other form may intervene between that form and the pronominal element.

One other piece of evidence suggests that the pronominal forms are clitics and not independent words. In Bontok and all other Central Cordilleran languages, any lexical item which is clearly independent, such as a personal or a common noun, and which can occur as the possessor of a noun or the Agent of a transitive clause, requires that the head of the construction, if otherwise ending in a vowel, end in *-n*. Compare examples (6–7), in which the head of each construction ends in a consonant, with (8–9), in which the head of each construction ends in a vowel plus *-n*. However, as shown in (10–11), the pronominal form following a head ending in a vowel may not have *-n* preceding it.

- (6) a. *ʔabuŋ Pakúlan* ‘Pakoran’s house’  
 b. *ʔabuŋ nan laláki* ‘the man’s house’
- (7) a. *ʔnwasít Pakúlan* ‘Pakoran threw (s.th.) out’  
 b. *ʔnwasít nan laláki* ‘the man threw (s.th.) out’
- (8) a. *ʔásu-n Pakúlan* ‘Pakoran’s dog.’  
 b. *ʔásu-n nan laláki* ‘the man’s dog’
- (9) a. *ʔníla-n Pakúlan* ‘Pakoran saw (s.th.)’  
 b. *ʔníla-n nan laláki* ‘the man saw (s.th.)’
- (10) a. *ʔabuŋ =yú* ‘your (PL) house’  
 b. *ʔásu =yú* ‘your (PL) dog’  
 c. *\*ʔásu-n yú*

<sup>8</sup> Brainerd labels these forms “ergative/genitive”, noting the homophony between not only the ergative and genitive pronominal forms, but also the homophony of her so-called “ergative/oblique case markers” and what she refers to as “genitive markers”, which she does not consider to be case forms (Brainerd 1997:145–146).

- (11) a. *ʔinwasít =yú* 'you (PL) threw (s.th.) out'  
 b. *ʔiníla =yú* 'you (PL) saw (s.th.)'  
 c. \*ʔiníla-n yú

## 2.2 Clitics or affixes?

Having determined that the forms are probably not free words is only part of the answer. Determining whether they are clitics or have lost their syntactic independence and become further grammaticalised as part of the word to which they were formerly phonologically attached is a more difficult task and has more intriguing theoretical implications.

Of the various criteria listed by Zwicky and Pullum (1983:503–504) and cited in A–F below, at least A, C and D are applicable to the forms under discussion.

- A. Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.  
 B. Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.  
 C. Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.  
 D. Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.  
 E. Syntactic rules can affect affixed words, but cannot affect clitic groups.  
 F. Clitics can attach to material already containing clitics, but affixes cannot.

With respect to A, there is a set of nonpronominal clitics in Bontok that do have a low degree of selection with respect to their hosts. This, for example, is true of all prepositions, such as locative *ʔas* and *ʔad*, either of which can become an enclitic, respectively =*s* and =*d*, to any vowel-final form that happens to precede them. It is also true of all determiners, which—depending on their form—become either preclitics or enclitics to the form to which they are adjacent. All of these clitics are typical of so-called “simple” clitics (Zwicky 1977), whose phonological attachment is clearly the result of their prosodically deficient character as monosyllabic forms and consequent loss of their vocalic nuclei, regardless of the fact that syntactically the prepositions and determiners are related to what follows them, not to what precedes them. In (12a), for example, the locative preposition *ʔad* is phonologically attached to a noun; in (12b), the same preposition is attached to a transitive verb (with a pronominal clitic) embedded in a relative clause; whereas in (12c), the preposition *ʔas* is attached to an infinitive, intransitive verbal complement. In this respect they are not unlike the NP-internal clitics of Kwakw’ala (Anderson 1997:9), which occur at the left edge of the phrase to which they are syntactically related, but which attach phonologically to the right edge of the preceding phrase.

- (12) a. *ʔinmáy nan laláki =d Mayníla*  
 went Det man =P Manila  
 ‘the man went to Manila’  
 b. *ʔinmáy nan laláki =ay ʔiníla =yú =d Mayníla*  
 went Det man =P saw =you =P Manila  
 ‘the man you saw went to Manila’

- c. *ñnmáy nan laláki =ay ñumála =s ñásu*  
 went Det man =P get =P dog  
 'the man went to get a dog'

The genitive pronominal forms, on the other hand, are more restricted in their distribution. They may attach only to nouns, or to transitive verbs that are the heads of 'root' clauses, that is, transitive verbs that are not themselves dependent on some other predicate. In this way then, they are similar to Zwicky's "special" clitics, and are more affixlike in their distribution than they are to 'simple' clitics. Their distribution as second-order clitics, either to the verbal head of a clause, or to the nominal head of a noun phrase is typical not only of many Austronesian languages, but of a wide range of languages from diverse language families, a fact noted first by Wackernagel (1892) and subsequently discussed in an extensive literature (Kaiser 1997).

With respect to C and D, there are several lines of evidence that suggest that, at least in Bontok, the shortened forms of the first and second singular pronominal forms exhibit morphophonological idiosyncrasies that make them more affixlike than cliticlike. The first is that these forms have not developed as the result of any regular phonological rule, otherwise we would expect that other pronominal forms ending in a high back vowel, such as =yú '2P', =takú '1+2P' and =kayú '2P', would also have shortened forms; but they do not.

Second, the strange fact that noncompleted transitive verbs replace their final *-n* with the postvocalic variant only in the first, second and third persons, but not for any other person and number, is an idiosyncrasy which makes the forms appear affixlike. When we consider other languages in the Cordilleran family that have short postvocalic pronominal forms, we find other idiosyncrasies that support the view that the shortened forms are not clitics but affixal. Both Yogad and Ibanag in the Cagayan Valley subgroup of Northern Cordilleran, as well as Itneg, Kalinga and Balangaw in the nuclear subgroup of Central Cordilleran have *-m* as the expected postvocalic form for '2S'. But instead of =*mu* as the postconsonantal form, each of these languages has =*nu*. The full clitic form has apparently evolved independently from the postvocalic variant, probably by analogy with the third person singular form which has an initial alveolar nasal, =*na* '3S'.<sup>9</sup>

In Ilokano, there are a series of morphophonological idiosyncrasies that are clearly of the kind that show that what were once genitive clitic pronouns have become incorporated into the verb and now function only to mark the verb as having certain agreement features. These idiosyncrasies occur in transitive verbs which imply a sequence of actor and Patient. When the sequence implies an actor followed by a third person nominative Patient, the forms are phonologically transparent only under certain conditions. When a third person singular pronoun (or any full lexical NP) occurs as the nominative Patient, then the expected form of the actor occurs—either the full, postconsonantal forms, as in Table 5a, or with the expected postvocalic forms if the transitive verb ends in a vowel. It should be noted that *isúna* '3S' is optional, and may occur independently, for instance as a nominal predicate.<sup>10</sup>

<sup>9</sup> Tharp (1974:86) reconstructs the Proto Northern Cordilleran second person singular form as \**nu*, *-m*. However, it seems clear that the source of the shortened form must originally have had a bilabial nasal, and that the change to an alveolar was a subsequent analogical development occurring in a cluster of geographically fairly closely related languages.

<sup>10</sup> Third person singular nominative pronouns are typically morphologically unmarked in many Philippine languages, as they are also in most Formosan languages.

**Table 5a:** Ilokano Transitive Actor agreement with 3S Nominative Patient

Actor Agreement	3S Nominative Patient Pronoun	
[+spkr,-addr,-plrl] '1S'	1. <i>kinábilkó (isúna)</i>	'I hit him/her'
[-spkr,+addr,-plrl] '2S'	2. <i>kinábilmó (isúna)</i>	'you (SG) hit him/her'
[-spkr,-addr,-plrl] '3S'	3. <i>kinábilná (isúna)</i>	'he hit him/her'
[+spkr,+addr,-plrl] '1+2S'	4. <i>kinábiltá (isúna)</i>	'we (DL) hit him/her'
[+spkr,-addr,+plrl] '1P'	5. <i>kinábilmí (isúna)</i>	'we (EX) hit him/her'
[+spkr,+addr,+plrl] '1+2P'	6. <i>kinábiltayó (isúna)</i>	'we (IN) hit him/her'
[-spkr,+addr,+plrl] '2P'	7. <i>kinábilyó (isúna)</i>	'you (PL) hit him/her'
[-spkr,-addr,+plrl] '3P'	8. <i>kinábilda (isuna)</i>	'they hit him/her'

A similar situation occurs when the nominative Patient is third person plural, as in Table 5b. However, in this case the form of the nominative clitic, although phonologically transparent, is not an independent lexical item. The independent predicative form is not *idá*, but *isúda* '3P'.

**Table 5b:** Ilokano Transitive Actor agreement with 3P Nominative Patient

Actor Agreement	3P Nominative Patient Pronoun Clitic	
[+spkr,-addr,-plrl] '1S'	9. <i>kinábilkó=idá</i>	'I hit them'
[-spkr,+addr,-plrl] '2S'	10. <i>kinábilmó=idá</i>	'you (SG) hit them'
[-spkr,-addr,-plrl] '3S'	11. <i>kinábilná=idá</i>	'he hit them'
[+spkr,+addr,-plrl] '1+2S'	12. <i>kinábiltá=idá</i>	'we (DL) hit them'
[+spkr,-addr,+plrl] '1P'	13. <i>kinábilmí=idá</i>	'we (EX) hit them'
[+spkr,+addr,+plrl] '1+2P'	14. <i>kinábiltayó=ida</i>	'we (IN) hit them'
[-spkr,+addr,+plrl] '2P'	15. <i>kinábilyó=idá</i>	'you (PL) hit them'
[-spkr,-addr,+plrl] '3P'	16. <i>kinábildá=idá</i>	'they hit them'

However, a different situation occurs where the form that marks the Patient is something other than third person. In this case only two possible forms can be identified as marking the transitive actor: one is *na*, as shown in Table 5c. But instead of marking a third person singular actor, *na* only indicates that the actor is singular. Person is not marked, hence the ambiguous interpretations for #17 and #20.



**Table 5c:** Ilokano [-spkr, -plrl] Transitive Actor agreement

Patient Agreement	Actor Agreement [-spkr,-plrl]	
[+spkr,-addr,-plrl] '1S'	17. <i>kinábilnák</i>	'you (SG) hit me' 'he hit me'
[-spkr,+addr,-plrl] '2S'	18. <i>kinábilnaká</i>	'he hit you (SG)'
[+spkr,+addr,-plrl] '1+2S'	19. <i>kinábilnatá</i>	'he hit us (DL)'
[+spkr,-addr,+plrl] '1P'	20. <i>kinábilnakamí</i>	'you (SG) hit us (EX)' 'he hit us (EX)'
[+spkr,+addr,+plrl] '1+2P'	21. <i>kinábilnatayó</i>	'he hit us (IN)'
[-spkr,+addr,+plrl] '2P'	22. <i>kinábilnakayó</i>	'he hit you (PL)'

The other possible form to identify the transitive actor is *da*, as shown in Table 5d. But *da* here only indicates that the actor is plural. Person is not marked hence the multiple ambiguities in the table.

**Table 5d:** Ilokano [+plrl] Transitive Actor agreement

Patient Agreement	Actor Agreement [+plrl]	
[+spkr,-addr,-plrl] '1S'	23. <i>kinábildák</i>	'you (PL) hit me' 'they hit me'
[-spkr,+addr,-plrl] '2S'	24. <i>kinábildaká</i>	'we (excl) hit you (SG)' 'they hit you (SG)'
[+spkr,+addr,-plrl] '1+2S'	25. <i>kinábildatá</i>	'they hit us (DL)'
[+spkr,-addr,+plrl] '1P'	26. <i>kinábildakamí</i>	'you (PL) hit us (EX)' 'they hit us (EX)'
[+spkr,+addr,+plrl] '1+2P'	27. <i>kinábildatayó</i>	'they hit us (IN)'
[-spkr,+addr,+plrl] '2P'	28. <i>kinábildakayó</i>	'we (EX) hit you (PL)' 'they hit you (PL)'

Finally, in Table 5e, there is no phonological form at all to mark the transitive actor. In this case the verbs are interpreted as having a first person singular actor.

**Table 5e:** Ilokano [+spkr,-plrl] Transitive Actor agreement

Patient Agreement	Actor Agreement [+spkr,-plrl]	
[-spkr,+addr,-plrl] '2S'	29. <i>kinábilká</i>	'I hit you (SG)'
[-spkr,+addr,+plrl] '2P'	30. <i>kinábilkayó</i>	'I hit you (PL)'

These changes that have taken place in Ilokano would suggest that when a sequence of clitics becomes incorporated into a word, integration happens one layer at the time, the earlier/innermost accretions becoming more phonologically integrated than the more recent accretions. Thus a sequence of what was once a genitive pronoun followed by a nominative



pronoun shows irregularity precisely in what was once the genitive form, but not in the nominative, which being at the outer boundary of the word has maintained its phonological transparency.

### 2.3 Affixes or agreement features?

Returning to the problem at hand, that of determining the nature of the so-called genitive pronouns, it is clear that even if the postconsonantal forms are treated as clitics, the postvocalic short forms cannot be considered to be phonologically conditioned “allomorphs” of them. All the evidence for the short forms clearly indicates that they are not clitics but are part of their earlier host, whether noun or verb. As clitics, the postconsonantal forms are pronouns, and in addition to having a case form—Genitive—they also carry a case relation, either Agent if they are attached to a transitive verb, or Correspondent if attached to a noun.<sup>11</sup> But as incorporated forms, they cannot by definition be pronouns, since they are now part of the noun or the verb that formerly hosted them. Nor can they have a case form, or a case relation. They have become simply forms that mark agreement, specifically with the person and number of the (optional) actor of transitive verbs, or of a Correspondent, if they are part of a noun.

There is some evidence that true clitics retain their phonological transparency, and therefore have the potential of moving away from their clitic status back to independent word, since this is primarily a phonological process with the only syntactic changes required being those that result from the loss of its clitic status. Such a change must have taken place in the history of languages (such as Inibaloi) which have lost an auxiliary verb (probably meaning ‘go’) that must have formerly marked imperatives, leaving the clitic pronouns which followed them stranded as independent forms.

- (13) Inibaloi (Ballard, Conrad & Longacre 1971:24)  
*Jo di olop jet idaw jo la'd ma Peshis.*  
 you here fetch and bring you *la*=to *ma* Peshis  
 ‘Go fetch him and bring him to Peshis.’

But once they have lost their clitic status and contributed their features to the host this is a process of no return. It is a one-way process with concomitant blurring of phonological boundaries triggered by the loss of morphological distinctiveness.

## 3 The analogical basis for the development of the Central Cordilleran Actor agreement marking

In the preceding sections I have used the term affix only as a convenient way of referring to forms that have been grammaticalised one step beyond their earlier status as clitics, and which in addition to becoming phonologically part of their host now contribute some of their semantic and syntactic features to it. At this point however, I shall abandon completely the

<sup>11</sup> A Correspondent is one of the five case relations allowed within Lexicase (Starosta 1988). It is, among others, the case expressed by a Genitive case form in Austronesian languages when it is dependent on a noun. In this position, it is typically referred to as a possessor. The other case relations are Agent, Patient, Locus and Means.

term 'affix', and refer simply to the agreement features that have been contributed to the head. This is an essential step to take. It is clear that retaining older structuralist notions of morphology, by which words are viewed as being composed of a root (or a stem) plus affixes, leads to unsolvable problems of segmentation in even 'analytic' languages such as English, with the postulating of replacive morphs and other such ad hoc devices. The problems become more acute in so-called 'fusional' languages, where it is often impossible to uniquely associate one or another syntactic or semantic feature with some phonological stuff within a word. The problem multiplies in complexity when one attempts to do a traditional morphological analysis of a Semitic language with its 'triconsonantal' roots.

More recent approaches to morphology (Starosta 1988, to appear; Anderson 1992; Ford, Singh & Martohardjono 1997) have moved beyond the structuralist Item and Arrangement (IA) and Item and Process (IP) attempts, and even earlier Word and Paradigm (WP) approaches, to so-called 'word-based', 'a-morphous', or 'seamless morphology' theories within which words are claimed not to have internal morphological structure.

### 3.1 Word formation strategies

Since words, according to the theory espoused here, specifically Lexicase as expounded by Starosta (1988 and subsequent works), have no internal morphological structure, they are themselves the smallest category available for syntactic analysis. There are no such things as 'morphology rules', that is rules which manipulate segmental morphemes. This does not, of course, mean that there is no relationship between the different forms of a verbal or nominal set. Words are related as members of analogical sets, formed by one or another WORD FORMATION STRATEGY (Ford, Singh & Martohardjono 1997:1). In Starosta's terms, a Word Formation Strategy "is an analogical pattern holding between pairs of words or n-tuples of words in a lexicon" (pers. comm.); in a grammar, it is a description of the relationship holding between a set of words that are perceived as sharing some analogical pattern. Thus, using Starosta's example (to appear, p. 5), the relationship between regular present and past tense English verbs is an analogical relationship of the form given in Figure 1, which can be represented by a word formation strategy of the type given in Figure 2.

slɪpt	:	slɪp	::
'slipped'		'slip'	
[+past]		[-past]	
mɪst	:	mɪs	::
'missed'		'miss'	
[+past]		[-past]	
wɔkt	:	wɔk	::
'walked'		'walk'	
[+past]		[-past]	

**Figure 1:** Analogical patterns: English *t* past and zero nonpast

[+past] : [-past]

t] : ]

**Figure 2:** Word formation strategy: English *t* past and zero nonpast

### 3.2 Bontok Word Formation Strategies

Applying this concept (of the analogical relationship between systematically related sets of words) enables us now to tackle the problem first stated in the introduction. There it was indicated that the primary purpose of this paper is to examine the factors that brought about the use of the postvocalic variants on transitive verbs that otherwise would end in a consonant. Restated without appealing to the concept of morphological variant, the issue is now to determine what the analogical processes were that resulted in the loss of the final *-n* of transitive verbs, when what were earlier genitive pronominal clitics became part of the verb and contributed their person and number features to the verb. The patterns first displayed in Table 3 are restated below in Table 6. I am taking the position here that the forms displayed in the first three rows are single lexical items—transitive verbs with Actor agreement features—whereas the forms in the remaining rows of the table are sequences of lexical items—transitive verbs followed by Genitive Agent clitic pronouns. This allows for a simpler description than one in which the forms in the remaining rows are also treated as marking actor agreement features on the verb. It is also consistent with the position that inflectional paradigms may develop one step at a time.<sup>12</sup>

**Table 6:** Bontok transitive verbs showing Actor agreement

1S	<i>ʔaláʔək</i>	'I get s.th.'	<i>dalusák</i>	'I clean s.th.'
2S	<i>ʔaláʔəm</i>	'you get s.th.'	<i>dalusám</i>	'you clean s.th.'
3S	<i>ʔaláʔəná</i>	'he/she gets s.th.'	<i>dalusáná</i>	'he/she cleans s.th.'
1+2S	<i>ʔaláʔən =tá</i>	'we (DL) get s.th.'	<i>dalusán =tá</i>	'we (DL) clean s.th.'
1P	<i>ʔaláʔən =mí</i>	'we (EX) get s.th.'	<i>dalusán =mí</i>	'we (EX) clean s.th.'
1+2P	<i>ʔaláʔən =takú</i>	'we (IN) get s.th.'	<i>dalusán =takú</i>	'we (IN) clean s.th.'
2P	<i>ʔaláʔən =yú</i>	'you (PL) get s.th.'	<i>dalusán =yú</i>	'you (PL) clean s.th.'
3P	<i>ʔaláʔən =dá</i>	'they get s.th.'	<i>dalusán =dá</i>	'they clean s.th.'

#### 3.2.1 The word formation strategy for vowel-final possessed nouns

The analogical patterns that are relevant to the problem include a set of nominal forms, specifically those that can be possessed and that end in a vowel. An example of such a set is given in Figure 3. The analogical word formation strategy (WFS1) that accounts for such forms is given immediately below the examples. Lexicase formalism is used to specify the relevant morphosyntactic features in the terms of the analogy. Included in Figure 3 are only those forms that are pertinent to the problem of understanding the basis of the analogical

<sup>12</sup> This is the position claimed by Starosta (1985) with reference to incipient case inflection in Mandarin Chinese.

association that brought about the irregular verbal forms under discussion. The set could have been extended to include all person and number forms, if they are considered to also be part of the noun and no longer clitics.

$\lambda\acute{a}su$ 'dog'	:	$\lambda\acute{a}suk$ 'my dog'	:	$\lambda\acute{a}sum$ 'your dog'	:	$\lambda\acute{a}sun$ 'dog of'	::
$b\acute{a}\eta a$ 'pot'	:	$b\acute{a}\eta ak$ 'my pot'	:	$b\acute{a}\eta am$ 'your pot'	:	$b\acute{a}\eta an$ 'pot of'	::
$sik\acute{i}$ 'leg'	:	$sik\acute{i}k$ 'my leg'	:	$sik\acute{i}m$ 'your leg'	:	$sik\acute{i}n$ 'leg of'	::
$\left[ \begin{array}{c} N \\ ?([N]) \\ \alpha F_i \end{array} \right]$	:	$\left[ \begin{array}{c} N \\ +pssd \\ ? \left[ \begin{array}{c} COR \\ +spkr \\ -addr \\ -plrl \end{array} \right] \\ ? [COR] \\ \alpha F_i \end{array} \right]$	:	$\left[ \begin{array}{c} N \\ +pssd \\ ? \left[ \begin{array}{c} COR \\ -spkr \\ +addr \\ -plrl \end{array} \right] \\ ? [COR] \\ \alpha F_i \end{array} \right]$	:	$\left[ \begin{array}{c} N \\ ? \left[ \begin{array}{c} N \\ -pmn \end{array} \right] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ -spkr \\ -addr \end{array} \right] \\ \alpha F_i \end{array} \right]$	
$Vj$		$Vk$		$Vm$		$Vn$	

Figure 3. WFS for Bontok possessed nouns (WFS1)

WFS1 states that there is an analogically based word formation strategy in Bontok such that any vowel-final noun ( $Vj$ ) which allows a dependent noun,  $?([N])$ —typically this would be a following 'possessor', or Correspondent in Lexicase terminology—may have associated with it a set of obligatorily possessed (+pssd) nouns each with the same set of semantic features ( $\alpha F_i$ ). One of these associated forms carries a set of agreement features which implies a first person singular possessor ( $[COR,+spkr,-addr,-plrl]$ ) and is marked with the ending  $k$ . Another of these associated forms carries a set of agreement features which implies a second person singular possessor ( $[COR,-spkr,+addr,-plrl]$ ) and is marked with the ending  $m$ . It should be emphasised that WFS1 encodes an assumption that at least the endings that mark first and second person agreement are *not* genitive clitics; the rule simply indicates that the whole form carries person and number pronominal features that agree with an (optional) dependent Correspondent. The endings are not themselves case-marked forms, nor are they suffixes in the traditionally understood use of the term.<sup>13</sup> A third associated form

<sup>13</sup> Typically in descriptions of Austronesian languages, affixed forms that are translatable as pronouns are referred to as pronouns, and are also noted as having a case form. Thus Zeitoun in discussing Budai Rukai nominative pronouns states, "The short form—the most widely used—occurs in postverbal position. It is suffixed to the verb, and its position is fixed" (1997:317).

requires a dependent Genitive nonpronominal noun (N, -prnn), which is a third person possessor ([COR, -spkr, -addr]). This form is marked with the ending *nj*. This is the form of a vowel-final noun as it appears before nonclitic (that is, nonpronominal) genitive noun phrases, as in (8) above, repeated below without any morphological divisions as (14):

- (14) a. *ʔásun Pakúlan* 'Pakoran's dog'  
 b. *ʔásun nan laláki* 'the man's dog'

### 3.2.2 The word formation strategy for vowel-final deverbal nouns

In the previous section, WFS1 showed an analogical relationship between nouns that allow a dependent Genitive Correspondent. This excludes certain nouns that are so-called "deverbal nominalizations". While for nonderived nouns such as *ʔásu* 'dog', *báŋa* 'pot' and *sikí* 'leg' a Genitive Correspondent is optional, a noun which has a derivational relationship with a vowel-final transitive verb must have either an agreement feature marking a possessor, as in (15a–b), or a dependent Genitive Correspondent—either a clitic pronoun, as in (15c), or a nonpronominal lexical noun, as in (15d–e). The noun *ʔníla* 'thing seen' is such a form.

- (15) a. *ʔásu nan ʔnílak* 'a dog is what I saw'  
 dog Det saw.of.1S lit. 'a dog is my seen thing'
- b. *ʔásu nan ʔnílam* 'a dog is what you (SG) saw'  
 dog Det saw.of.2S lit. 'a dog is your (SG) seen thing'
- c. *ʔásu nan ʔníla =yu* 'a dog is what you (PL) saw'  
 dog Det saw.of =Gen.2P lit. 'a dog is your (PL) seen thing'
- d. *ʔásu nan ʔnílan Pakúlan* 'a dog is what Pakoran saw'  
 dog Det saw.of Pakoran lit. 'a dog is the seen thing of Pakoran'
- e. *ʔásu nan ʔnílan nan laláki* 'a dog is what the man saw'  
 dog Det saw.of Det man lit. 'a dog is the seen thing of the man'

The word formation strategy that relates sets such as this is given in Figure 4.

<i>ʔn̄ila</i>	<i>ʔn̄ilak</i>	<i>ʔn̄ilam</i>	<i>ʔn̄ilan</i>	::
'seen by prnn'	'seen by me'	'seen by you'	'seen by N'	
<i>ʔnláku</i>	<i>ʔnlákuk</i>	<i>ʔnlákum</i>	<i>ʔnlákun</i>	::
'sold by prnn'	'sold by me'	'sold by you'	'sold by N'	
<i>ʔláku</i>	<i>ʔlákuk</i>	<i>ʔlákum</i>	<i>ʔlákun</i>	::
'sells by prnn'	'sells by me'	'sells by you'	'sells by N'	
$\left[ \begin{array}{c} N \\ +pssd \\ ?[N] \\ ? \left[ \begin{array}{c} COR \\ +prnn \\ +cltc \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ +Gen \end{array} \right] \\ ?[COR] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} N \\ +pssd \\ ? \left[ \begin{array}{c} COR \\ +spkr \\ -adrs \\ -plr \end{array} \right] \\ ?[COR] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} N \\ +pssd \\ ? \left[ \begin{array}{c} COR \\ -spkr \\ +adrs \\ -plr \end{array} \right] \\ ?[COR] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} N \\ +pssd \\ ?[N] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ -spkr \\ -adrs \end{array} \right] \\ \alpha F_i \end{array} \right]$	
<i>Vj</i>	<i>Vkj</i>	<i>Vmj</i>	<i>Vnj</i>	

Figure 4: WFS for Bontok nominalised vowel-final transitive verbs (WFS2)

WFS2 states that there is an analogically based word formation strategy in Bontok such that any vowel-final (*Vj*) possessed (+pssd) noun requiring a pronominal Genitive clitic dependent may have associated with it a set of possessed nouns each with the same set of semantic features ( $\alpha F_i$ ). One of these associated forms implies a first person singular possessor and is marked with the ending *kj*. Another of these associated forms implies a second person singular possessor and is marked with the ending *mj*. A third associated form requires a dependent Genitive nonpronominal noun (N, -prnn), which is a third person possessor and is marked with the ending *nj*.

3.2.3 *The word formation strategy for vowel-final transitive verbs*

Vowel-final transitive verbs (and the nouns with which they are analogically related) are of three types. These are exemplified at the head of Figures 4 and 5. The first are vowel-final verbs that have an initial [*ʔC* sequence, where the *C* is not an alveolar nasal consonant. These have traditionally been called nonperfective instrument (or associate) focus verbs with an *ʔn̄*- prefix, such as *ʔláku* 'sell s.th.', *ʔpili* 'choose s.th.', *ʔyáli* 'bring s.th.'. The second are the perfective forms of the same verbs. In Bontok, these verbs have an initial [*ʔn̄* sequence,<sup>14</sup>

<sup>14</sup> Historically these initial sequences resulted from what has in the past been considered to be affixation of a perfective <in> into an *ʔn̄*- prefixed stem, resulting in a sequence of *ʔn̄in>i-*, with subsequent deletion of the final (unstressed) *i* of the affix complex.

such as *ʔnɪ́laku* ‘sold s.th.’, *ʔnpɪ́li* ‘chose s.th.’, *ʔnyá́li* ‘brought s.th.’. The third type of vowel-final transitive verbs are the perfective forms of so-called ‘object (or goal) focus’ verbs with an initial *[Cin]* sequence. All other transitive verbs are consonant-final, since they end in reflexes of \*-an or \*-en. The word formation strategy that relates these verbs is given in Figure 5.

<i>ʔnɪ́la</i> ‘prmn saw s.th.’	<i>ʔnɪ́lak</i> ‘I saw s.th.’	<i>ʔnɪ́lam</i> ‘you saw s.th.’	<i>ʔnɪ́lan</i> ‘N saw s.th.’
<i>ʔnláku</i> ‘prmn sold s.th.’	<i>ʔnlákuk</i> ‘I sold s.th.’	<i>ʔnlákum</i> ‘you sold s.th.’	<i>ʔnlákun</i> ‘N sold s.th.’
<i>ʔlákua</i> ‘prmn sells s.th.’	<i>ʔlákuk</i> ‘I sell s.th.’	<i>ʔlákum</i> ‘you sell s.th.’	<i>ʔlákun</i> ‘N sells s.th.’

$\left[ \begin{array}{c} V \\ +trns \\ ? [N] \\ ? \left[ \begin{array}{c} +Gen \\ +prnn \\ +cltc \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ +Gen \end{array} \right] \\ ?[AGT] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? \left[ \begin{array}{c} AGT \\ +spkr \\ -adrs \\ -plrl \end{array} \right] \\ ? [AGT] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ +adrs \\ -plrl \end{array} \right] \\ ?[AGT] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? [N] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ -adrs \end{array} \right] \\ \alpha F_i \end{array} \right]$
<i>Vj</i>	<i>Vkj</i>	<i>Vmj</i>	<i>Vnj</i>

Figure 5: WFS for Bontok vowel-final transitive verbs (WFS3)

WFS3 states that there is an analogically based word formation strategy in Bontok such that any vowel-final transitive verb (*Vj*) which requires a pronominal Genitive clitic dependent expressing an Agent, may have associated with it a set of transitive verbs each with the same set of semantic features ( $\alpha F_i$ ). One of these associated forms implies a first person singular actor and is marked with the ending *kj*. Another of these associated forms implies a second person singular actor and is marked with the ending *mj*. A third associated form requires a dependent Genitive nonpronominal noun (N, -prnn), which is a third person Agent and is marked with the ending *nj*.

### 3.2.4 The word formation strategy relating vowel-final transitive verbs and their nominalisations

Vowel-final transitive verbs and their homophonous nominalised forms are themselves related by a word formation strategy, two terms of which are given in Figure 6.

<i>ʔnılan</i> 'N saw s.th.'	:	<i>ʔnılan</i> 'thing seen by N'
<i>ʔnlákun</i> 'N sold s.th.'	:	<i>ʔnlákun</i> 'thing sold by N'
<i>ʔlákun</i> 'N sells s.th.'	:	<i>ʔlákun</i> 'thing being sold by N'
$\left[ \begin{array}{c} V \\ +trns \\ ? [N] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ -adrs \end{array} \right] \\ \alpha F_i \end{array} \right]$	:	$\left[ \begin{array}{c} N \\ +pssd \\ ? [N] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} COR \\ -spkr \\ -adrs \end{array} \right] \\ \alpha F_j \end{array} \right]$
<i>n]</i>		<i>n]</i>

**Figure 6:** WFS for Bontok transitive verbs to nominalised perfective transitive verbs (WFS4)

WFS3 states that there is an analogically based word formation strategy in Bontok such that any verb which has a Genitive Agent, third person lexical noun (i.e. nonpronominal) dependent has an analogical relationship with a possessed noun of the same form. Both forms end in *n]*, but whereas the case relation carried by the noun that is the Genitive dependent of the verb is Agent, that of the associated noun is Correspondent, or possessor. The verb in this situation would be, by definition within Lexicase (Starosta 1988), a transitive verb, since only transitive verbs may have an Agent.

### 3.2.5 *The Word Formation Strategy for transitive verbs ending in n]*

The analogical pattern existing between sets of vowel-final nouns and verbs, by which the addition of an *n]* ending marks the presence of a Genitive nonpronominal dependent, provided the basis for the innovation first discussed in §1, and illustrated in Tables 3 and 6, by which so-called 'goal focus' transitive verbs containing a reflex of PPh \*-en, and 'locative focus' transitive verbs containing a reflex of PPh \*-an (as well as their nominal counterparts) lose the final *n]* of the form when it carries first, second or third person singular agreement marking. In effect, forms which are consonant-final are treated as though they are vowel-final, their final *n]* marking the presence of a Genitive third person lexical noun dependent, as shown in Figure 7. A parallel word formation strategy exists relating the homophonous nominalised forms of these transitive verbs.



<i>ʔlákun</i> : 'N sells s.th.'	<i>ʔlákuk</i> : 'I sell s.th.'	<i>ʔlákum</i> : 'you sell s.th.'	<i>ʔlákuna</i> :: 'he sells s.th.'
<i>ʔlápən</i> : 'N sees s.th..'	<i>ʔlápək</i> : 'I see s.th.'	<i>ʔlápəm</i> : 'you see s.th.'	<i>ʔlápəna</i> :: 'he sees s.th.'
<i>dalusán</i> : 'N cleans s.th.'	<i>dalusák</i> : 'I clean s.th.'	<i>dalusám</i> : 'you clean s.th.'	<i>dalusána</i> :: 'he cleans s.th.'

$\left[ \begin{array}{c} V \\ +trns \\ ? [N] \\ ? \left[ \begin{array}{c} +Gen \\ N \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ +Gen \end{array} \right] \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ -adrs \end{array} \right] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? \left[ \begin{array}{c} AGT \\ +spkr \\ -adrs \\ -plrl \end{array} \right] \\ ? [AGT] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ +adrs \\ -plrl \end{array} \right] \\ ? [AGT] \\ \alpha F_i \end{array} \right]$	$\left[ \begin{array}{c} V \\ +trns \\ ? \left[ \begin{array}{c} AGT \\ -spkr \\ -adrs \\ -plrl \end{array} \right] \\ ? [AGT] \\ \alpha F_i \end{array} \right]$
<i>Vn/</i>	<i>Vk/</i>	<i>Vm/</i>	<i>Vna/</i>

Figure 7: WFS for *n/*- final transitive verbs (WFS5)

#### 4 Historical source of *n/*

The analogically based word formation strategies labelled WFS1–WFS5 are common to all Central Cordilleran languages and must therefore be reconstructable to their parent language. It is clear that the sources of the *k/* and *m/* first and second person singular agreement markers on postvocalic forms are the reduced forms respectively of the earlier clitics =*ku* and =*mu*, so that the Genitive case-marking and agreement system of Proto Central Cordilleran can be represented as in Table 7:

Table 7: Proto Central Cordilleran Genitive case-marking and agreement system

	POSTCONSONANTAL	POSTVOCALIC
Pronoun: 1S	= <i>ku</i>	<i>k/</i>
2S	= <i>mu</i>	<i>m/</i>
Proper Noun	N	<i>n/</i> N
Common Noun	Det N	<i>n/</i> Det N

However, a question arises as to the source of the final *n/* on postvocalic forms. It seems probable that it is the result of the grammaticalisation of what was earlier a genitive case-marking preposition \**ni* which occurred postvocally in Proto Cordilleran and earlier stages

of the family (Reid 1981), as shown in Table 8. Loss of the postconsonantal form \*ʔi with subsequent cliticisation of postvocalic \*ni onto vowel-final stems, and finally, with loss of its vocalic nucleus, complete phonological merger and bequeathing of its features to its head resulted in the *n*] final forms.

**Table 8:** Proto Cordilleran Genitive marking

	POSTCONSONANTAL	POSTVOCALIC
Pronoun: 1s	= <i>ku</i>	= <i>ku</i>
2s	= <i>mu</i>	= <i>mu</i>
Proper Noun	<i>ʔ</i> N	<i>ni</i> N
Common Noun	<i>ʔ</i> Det N	<i>ni</i> Det N

## 5 Conclusion

This study has attempted to provide an explanation for the unusual verbal and nominal forms in the Central Cordilleran languages, in which what were originally derivational endings, *en*] and *an*], lose their final nasal when they carry either first, second or third person singular agreement features, respectively *k*], *m*] and *na*]. I have claimed that this situation arose as a result of the incorporation of reduced forms of what were originally clitic genitive pronouns on the one hand, and subsequently the incorporation of a reduced form of the genitive case-marking preposition *ni* as *n*] on the other hand, onto vowel-final verbs and their nominal counterparts. An analogy was then established between these *n*] final forms and derived forms with *en*] and *an*] endings, so that the latter were also perceived as being vowel-final for the purpose of substituting the *k*], *m*] and *na*] agreement endings.

This explanation, however, does not account for the presence of a similar phenomenon in Ilokano (Rubino 1997), a language which does not form a part of the Central Cordilleran subgroup and which has been considered to be a first-order subgroup within the Cordilleran family. The explanation does not account for the Ilokano situation because this language does not have, and has no evidence of ever having had, a reduced form of the genitive *ni* occurring on vowel-final forms, a situation which I have claimed was crucial to the development of the agreement markers in the Central Cordilleran languages. Ilokano does have a set of *n* initial clitics that attach to vowel-final words. These are =*nto* 'future' and =*nsa* 'maybe, possible'. Alternate forms without the initial nasal attach to consonant-final words. Similarly, Ilokano =*n* 'now, already' attaches to vowel-final words, while its alternate, =*en* attaches to consonant-final words. There is no evidence, however, that the nasal on these forms was originally \**ni*, nor that they formed the basis of an analogy such as we find in the Central Cordilleran languages.

A number of possible explanations exist for the Ilokano situation, none of which I find particularly appealing. One explanation is that Ilokano has been influenced by its geographical proximity to the Central Cordilleran languages, and has developed precisely the same set of agreement features by a process of borrowing. I find this unconvincing, because Ilokano is, and probably has been for centuries, a more prestigious language than any of the Central Cordilleran languages, and lexical borrowing has been almost completely in the opposite direction.

A second explanation is that we don't know enough of the historical development of the Ilokano case-marking system. The conditions that brought about the agreement marking

system in the Central Cordilleran languages may well have existed in Ilokano at some earlier point in its history. There has been a substantial reformation in Ilokano of the system reconstructed for Proto Cordilleran, as can be seen by comparing Tables 8 and 9. The distinction between postconsonantal and postvocalic forms has been lost in Ilokano, too, with *ni* being generalised to follow both consonants and vowels, but only preceding proper nouns. Common nouns are preceded by *ti* or one of a variety of other forms having their origins in demonstratives. Although included in Table 9, neither *ni* nor *ti* are case markers since they also precede Nominative nouns and other case forms. Most cases in Ilokano are marked by word order and by the forms of pronouns. I find this second explanation unappealing because there is no alternation present in the language between a final *n]* and the form *ni*, as occurs, for example, between *n]* and *no* in the phrase *inton bigat* or *into no bigat* 'tomorrow'. Moreover, the *n-* initial clitics that do occur in Ilokano are very ancient, predating the phonological reduction of first and second person singular pronominal forms *ko* and *mo*. When *n-* initial clitics occur following these pronominal forms, they protect them from phonological reduction, as in *nakitamon* 'you saw (it) already'.

Table 9: Ilokano Genitive marking

	POSTCONSONANTAL	POSTVOCALIC
Pronoun: 1S	= <i>ku</i>	<i>k]</i>
2S	= <i>mu</i>	<i>m]</i>
Proper Noun	<i>ni</i> N	<i>ni</i> N
Common Noun	<i>ti</i> N	<i>ti</i> N

A third explanation is that Ilokano is actually a Central Cordilleran language and shares the innovation with those languages. But this would create more problems because Ilokano does not share any of the innovations which link both the Central and Southern Cordilleran languages, and would still require an explanation for the loss of word-final *n]* from *ni*.

Finally, it may be that my account of the development of agreement markers in the Central Cordilleran languages is wrong, or that there are other as yet unrecognised explanations for the irregular loss of the final nasal of transitive verbs that have operated in Ilokano. Zeitoun (1997:327) discusses a similar situation also in Kavalan, in which the final nasal of an *-an* suffix is lost before what appear to be *k* '1S' and *na* '3S' agreement markers. She describes these, in the traditional manner, as Genitive bound pronouns, as in (16a & b) (cited with her glosses<sup>15</sup>). She does not discuss the conditions that may have brought about this situation.

- (16) a. *pukun -an -ku -pa sunis* [*pukunaka sunis*]  
 beat -P/LF -1S.GEN -will child  
 'I will beat the child'
- b. *pukun -an -na* [*pukunana*]  
 beat -P/LF -3S.GEN  
 'he beat (s.o.)'

Perhaps when the conditions for the development of the Kavalan agreement markers become clear, a more comprehensive explanation will be available for both Ilokano and the languages of the Central Cordillera.

<sup>15</sup> P/LF = Patient or Locative Focus

## References

- Anderson, Stephen R., 1992, *A-morphous morphology*. Cambridge: Cambridge University Press.
- 1997, Towards an optimal account of second position phenomena. In Lizanne Kaiser, ed. *Yale A-morphous linguistic essays: studies in the morphosyntax of clitics*, 1–27. New Haven: Department of Linguistics, Yale University.
- Ballard, Lee D., Robert J. Conrad and Robert B. Longacre, 1971, *More on the deep and surface grammar of interclausal relations*. Language Data: Asian-Pacific Series 1. Santa Ana, California: Summer Institute of Linguistics.
- Blust, Robert A., 1977, The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9(2):1–16. Honolulu: Department of Linguistics, University of Hawai‘i.
- Brainerd, Sherri, 1997, Ergativity and grammatical relations in Karao. In T. Givon, ed. *Grammatical relations: a functionalist perspective*, vol. 35, Typological Studies in Language, 85–154. Philadelphia: John Benjamins.
- Dyen, Isidore, 1974, The Proto-Austronesian enclitic genitive pronouns. Proceedings of the First International Conference on Comparative Austronesian Linguistics, vol. 2: Proto-Austronesian and Western Austronesian. *Oceanic Linguistics* 13:17–31.
- Ford, Alan, Rajendra Singh and Gita Martohardjono, 1997, *Pace Pāṇini*. American University Studies, series XIII, vol. 34. New York: Peter Lang.
- Headland, Thomas N. and Janet D. Headland, 1974, *A Dumagat (Casiguran)-dictionary*. Canberra: Pacific Linguistics.
- Healey, Alan, 1958, Notes on Yogad. *Oceania Linguistic Monographs* 3:77–82.
- Kaiser, Lizanne, ed., 1997, *Yale A-morphous linguistic essays: studies in the morphosyntax of clitics*. New Haven: Department of Linguistics, Yale University.
- Reid, Lawrence A., 1974, The Central Cordilleran subgroup of Philippine languages. *Oceanic Linguistics* 13:511–560.
- 1979, Towards a reconstruction of the pronominal systems of Proto-Cordilleran, Philippines. In Nguyen Dang Liem, ed. *South-East Asian Linguistic Studies*, vol. 3:259–275. Canberra: Pacific Linguistics.
- 1981, Proto-Austronesian genitive determiners. In Andrew Gonzalez and David Thomas, eds *Linguistics Across Continents, Studies in Honor of Richard Pittman*. Linguistic Society of the Philippines Monograph Series, No. 11:97–105.
- 1983, Fieldnotes on Dupanangan (Palanan) Agta and Malaweg.
- 1989, Arta, another Philippine Negrito language. *Oceanic Linguistics* 28(1):47–74.
- 1991, The Alta languages of the Philippines. In Ray Harlow, ed. *VICAL 2: Western Austronesian and Contact Languages*, 265–297. Papers from the Fifth International Conference on Austronesian Linguistics. Te Reo Special Publication. Linguistic Society of New Zealand.

- Rubino, Carl R.G., 1997, A reference grammar of Ilocano. PhD dissertation, University of California, Santa Barbara.
- Starosta, Stanley, 1985, Mandarin case marking: a localistic lexicase analysis. *Journal of Chinese Linguistics* 13.2:215–266.
- 1988, *The case for lexicase: an outline of lexicase grammatical theory*. Open Linguistic Series. London: Pinter Publishers.
- to appear. Do compounds have internal structure?
- Tharp, James A., 1974, The Northern Cordilleran subgroup of Philippine languages. *Working Papers in Linguistics* 6(6):53–114. Honolulu: Department of Linguistics, University of Hawai'i.
- Wackernagel, Jacob, 1892, Über ein Gesetz der indogermanischen Wortstellung. *Indogermanisch Forschungen* 1:333–436.
- Zeitoun, Elizabeth, 1997, The pronominal system of Mantauran (Rukai). *Oceanic Linguistics* 36(2):312–346.
- Zwicky, Arnold M., 1977, *On clitics*. Bloomington: Indiana University Linguistics Club.
- 1985, Clitics and particles. *Language* 61(2):283–305.
- Zwicky, Arnold M. and Geoffrey K. Pullum, 1983, Cliticization vs. inflection: English *n't*. *Language* 59(3):502–513.



# *Proto Oceanic \*i, \*qi, and \*-ki*

---

MALCOLM D. ROSS

Hooper (1985) reconstructs POc \*qi, a particle that occurred between an inalienably possessed noun and its nonspecific possessor. However, she notes some irregularities in the reflexes of this morpheme, as POc \*q is sometimes reflected as if it were \*k. Hooper also reconstructs a personal article \*i that occurs in noun phrases where the possessor is a personal noun phrase, and alludes to the POc locative preposition \*(q)i reconstructed by Pawley (1972:85). My examination of a range of functional and phonological evidence concludes that these morphemes and others cognate with them reflect four POc morphemes: a personal article \*i; a nonspecific inalienable possessive marker \*qi; a free-form derivative suffix \*-ki; and a locative preposition \*i.

## 1 Introduction<sup>1</sup>

This short paper is an attempt to deal with a detail of Proto Oceanic (POc) morphology.<sup>2</sup> In a careful and well-argued paper, Hooper (1985) reconstructs POc \*qi, a particle that occurred between an inalienably possessed noun and its nonspecific possessor in the construction exemplified in (1).<sup>3</sup>

---

<sup>1</sup> This paper is a by-product of a project on the reconstruction of Proto Oceanic adjectives, their morphosyntax and their history, which Byron Bender, in his role as editor of *Oceanic Linguistics*, encouraged me to write up as Ross (1998a) and Ross (1998b). It is a token of my gratitude for Byron's continuing encouragement, as well as for the contribution he makes to Oceanic linguistics through his proactive and committed administration of the field's flagship journal. With regard to the current paper, I am indebted to Andrew Pawley and John Lynch for their comments and for drawing my attention to data I would otherwise have missed, and I owe special thanks to Catriona Hyslop, whose North-East Ambae data first sparked my interest in unpossessed forms of inalienable nouns. Without the data she generously supplied, this paper would not exist.

<sup>2</sup> Abbreviations for person and number are formatted thus: 1 – first person, 2 – second person, 3 – third person, D – dual, E – exclusive, I – inclusive, P – plural, S – singular; D: – disjunctive (=free), S: – subject, and P: – possessor, are prefixed to possessor pronoun forms (e.g. D:3D – third person dual disjunctive pronoun). Other abbreviations used in glosses are: ART – article, CL – possessive classifier, FREE – free-form derivative suffix, LIG – ligature, NCL – numeral classifier, PREP – preposition.

<sup>3</sup> Reconstructions are mine. In Ross (1998a) and in an earlier version of the present paper I reconstructed \*qi as an enclitic. Although there is good reason to suppose a degree of phonological bondedness between the possessed noun and \*qi, I have abandoned the enclitic interpretation for reasons given in Ross (1998b).

---

Joel Bradshaw and Kenneth L. Rehg, eds, *Issues in Austronesian morphology: a festschrift for Byron W. Bender*, 259–278.

Canberra: Pacific Linguistics 2001

© Malcolm D. Ross

Ross, M.D. 'Proto Oceanic \*i, \*q, and \*-ki'. In Bradshaw, J. and Rehg, K. editors, *Issues in Austronesian Morphology: A festschrift for Byron W. Bender*. PL-519-259-278. Pacific Linguistics, The Australian National University, 2001. DOI:10.1344/PL-519.259

©2001 Pacific Linguistics and/or the author(s). Online edition licensed 2015 CC BY-SA 4.0, with permission of PL. [Asealang.net/CRCL](http://Asealang.net/CRCL) initiative.

- (1) POC  
 \*a qaqe qi boRok  
 ART leg qi pig  
 'a pig's leg, leg of pork'

She also remarks on the fact that there are some irregularities in the reflexes of this morpheme, as POC \*q is sometimes reflected as if it were \*k. This is true particularly of those North Vanuatu reflexes that appear to be suffixed to an inalienable noun to produce an unpossessed (= independent) form, such as North-East Ambae *vulu-ki* 'hair (possessor not mentioned)', *vulu-gu* 'my hair' (Catriona Hyslop, pers. comm.). Hooper also reconstructs a personal article \*i that occurs in noun phrases where the possessor is a personal noun phrase (see below), and alludes to the POC locative preposition \*(q)i reconstructed by Pawley (1972:85).

The question I am addressing here is: how many Proto Oceanic morphemes do these reflexes represent? I will attempt to answer this question by looking at both functional and phonological evidence.

## 2 The Proto Oceanic possession system

The construction in (1) contrasted with one in which the possessor was specific:

- (2) POC  
 \*a qaqe-ña tam<sup>w</sup>ata  
 ART leg-P:3S man  
 'a/the leg of a certain/the man'

We can be reasonably certain from the work of Lichtenberk (1985) and Hooper (1985) (and others before and since, including Pawley 1972, 1973; Pawley & Sayaba 1990; Lynch 1982, 1996a, 1996b) that the Proto Oceanic possessive system was as set out in Table 1.

As Table 1 shows, possession constructions in Proto Oceanic varied along two parameters. First, the possessed noun belonged to one of two morphological classes, inalienable or free. Members of the inalienable class were mostly kin terms and parts of wholes, but class membership appears to have been lexically determined: in modern Oceanic languages, one cannot predict with certainty from their semantics which nouns will be inalienable. The free class was the default: it contained all nouns not in the inalienable class.

The second parameter concerned the status of the possessor noun phrase. Table 1 shows a two-way distinction between a specific and a nonspecific possessor. A nonspecific noun denoted a class or a class member, but not a particular member that the speaker wished to refer to. Semantically, nonspecific 'possessors' were often not really possessors at all, but generic nouns used attributively, as (1) shows. A nonspecific possessor had by definition to be a noun.



**Table 1:** Proto Oceanic noun phrases with common noun phrase possessors\*  
(after Hooper 1985 and Lichtenberk 1985)

POSSESSOR	POSSESSED	
	INALIENABLE	FREE
SPECIFIC		
PERSONAL	ART D-ART R *a qaqe-i X ART leg-ART X 'X's leg'	ART D CL-ART R *a Rumaq na-i X ART house CL-ART X 'X's house'
PRONOMINAL	D-P: *a qaqe-gu ART leg-P:1S 'my leg'	D CL-P: *a na-gu Rumaq ART CL-P:1S house 'my house'
COMMON	D-P:R *a qaqe-ña tam <sup>w</sup> ata ART leg-P:3S man 'the man's leg'	D CL-P:R *a na-ña Rumaq tam <sup>w</sup> ata ART CL-P:3S house man 'the man's house'
NONSPECIFIC (COMMON)	D qi R *a natu qi boRok ART child qi pig 'a piglet' (lit. 'child of pig')	D ni R *a polo ni niuR ART liquid ni coconut 'coconut water'

\* The abbreviations R and D are used here for the possessor and possessed noun phrases respectively.

Specific possessors are divided on formal grounds into personal, pronominal and common. If the possessor was pronominal, then its person and number were marked by a possessor suffix which was attached directly to an inalienable possessed noun or, if the possessed noun was free, to a possessive classifier, in this case the general classifier \*na-<sup>4</sup>:

- (3) POC
- |            |                |                |                      |
|------------|----------------|----------------|----------------------|
| *a qaqe-gu | 'my leg'       | *a na-gu Rumaq | 'my house'           |
| *a qaqe-mu | 'your (S) leg' | *a na-mu Rumaq | 'your (S) house'     |
| *a qaqe-ña | 'her/his leg'  | *a na-ña Rumaq | 'her/his house' etc. |

<sup>4</sup> The possessive classifier \*na-, the general or default classifier, is one of three reconstructed by Lichtenberk (1985); the others are \*ka- 'food' and \*ma- 'drink'. Questions about how many classifiers there were in POC and their forms and uses remain under discussion (see, for example, Lynch 1996b), but the details of their reconstruction have no bearing on the topic of this paper. It is possible that the default classifier (\*na-) was not preceded by the article (indeed, historically it may have been the same morpheme) but that other classifiers were. I have not adopted this possibility in the reconstructions here.

When the specific possessor was a common noun phrase, this was simply appended to the phrase, as shown in Table 1.<sup>5</sup> When, however, it was a personal noun phrase (one in which the head was a proper noun or perhaps a kin term), it appears that its personal article \*i was cliticised directly to the possessed noun or classifier, as in:

- (4) POC  
 \*a qaqe=i X  
 ART leg=ART X  
 'X's leg'

Hooper (1985:157) suggests that the possessor suffix was also present with a personal possessor, as it was with a common possessor. That is, she would reconstruct (5) rather than (4):

- (5) POC  
 \*a qaqe-ña i X  
 ART leg-P:3S ART X  
 'X's leg'

Certainly there are languages where this structure is reflected. Hooper gives examples from a number of languages, but among these the sequence of possessor suffix and personal article occurs only in Tigak and Roviana (it is also found in Arosi). However, there are also widely distributed languages—widely distributed, that is, both geographically and genetically—which, like (6b), reflect the structure in (4).<sup>6</sup>

- (6) BOUMAA FIJIAN
- a. *a liŋa-gu*  
 ART hand-P:1S  
 'my hand'
  - b. *a liŋa-i Jone*  
 ART hand-i John  
 'John's hand'
  - c. *a liŋa-na a noneyalewa yai*  
 ART hand-P:3S ART young.girl this  
 'this young girl's hand'

Clearly, analogical regularisation has occurred in a good many daughter-languages. It has gone in one of two directions. If the reconstruction in Table 1 is correct and (4) was the Proto Oceanic structure, then in Tigak, Roviana, Arosi and many other languages apparently reflecting the personal possessor construction in (5), the personal possessor construction has been analogically 'infected' by the common. If, on the other hand, (5) is the correct reconstruction, then no analogical infection has occurred.

Again assuming the reconstructions in Table 1 to be correct, there are languages, among them Tolai, North-East Ambae and Anejoñ, where the infection has perhaps worked in the

<sup>5</sup> For the free construction with a common noun phrase possessor, both \*a na-ña Rumaq tam<sup>w</sup>ata and \*a Rumaq na-ña tam<sup>w</sup>ata 'the man's house' are reconstructable. I suspect that, as in some modern Oceanic languages, the two orders coexisted, the difference between them being one of information structure.

<sup>6</sup> My assumptions about Oceanic subgrouping for reconstructive purposes and the grounds for them are given in Ross (1998a).

opposite direction, and the common possessor construction seems to have been infected by the personal. In these languages, the reflex of \*i has lost its general personal article function and occurs only in possession phrases. As a result it has been reinterpreted as a possessive connective. This has evidently started to happen in Boumaa Fijian, where (7) is an alternative to (6c) (Dixon 1988:123). This development is easier to explain if we assume that (4) was the Proto Oceanic structure.

- (7) BOUMAA FIJIAN  
*a lija-i ɲoneyalewa yai*  
 ART hand-P:3S young.girl this  
 'this young girl's hand'

Happily, this issue is also not particularly important to the present discussion, and it is possible that (4) and (5) co-occurred in the protolanguage.

### 3 Distinguishing between \*i 'personal article' and \*qi 'nonspecific possessive particle'

Hooper discusses the possibility that \*i in (4) or (5) was the same morpheme as her \*qi in (1) and concludes that it was not. Her reasons are compelling and are encapsulated in Table 1: \*i was selected by the (personal) possessor noun phrase, \*qi by the possessed (inalienable) noun.

I concur with Hooper. If \*i and \*qi had been the same morpheme, then we would have to reconstruct a general possessive preposition \*(q)i. We would then expect this \*(q)i to occur not only with specific personal possessors and with nonspecific (common) possessors, but also with specific common possessors. But its absence from the specific common possessor cells in Table 1 is well founded. In Table 2 are tabulated, among other morphemes and interpretive information discussed below, reflexes of \*i and \*qi in selected Oceanic languages according to their occurrence in specific personal, specific common and (nonspecific) common possessor constructions. (I explain below why a form like *na* should count as a reflex of \*qi.) In four widely separated languages, Tigak (Western Oceanic, Meso-Melanesian), Mota (North Vanuatu), Kwamera (South Vanuatu) and Bauan, and Boumaa Fijian (Central Pacific), we find a significant distribution: \*i is reflected with a specific personal possessor, \*qi with nonspecific possessor. In these languages, no morpheme reflecting either \*i or \*qi occurs in the specific common possessor construction. If the reflexes of \*i and \*qi were simply reflexes of a general possessive \*(q)i, this would be a very strange distribution indeed. But if, as Hooper argues, we are dealing with two distinct morphemes, there is no problem.

Table 2 shows a number of languages where a reflex of \*i or \*qi does occur in the specific common possessor construction, but this is not surprising. As I noted above, in a number of languages \*i has been reanalysed as a possessive connective and has extended (or is extending) its domain from specific personal to specific common possession. In Tamambo (North Vanuatu), the reflex of \*qi has apparently extended its domain from nonspecific possession, which is by definition common, to specific common possession. In languages like North-East Ambae, where the three categories of possession are marked in the same way, it is probably vacuous to ask whether the morpheme in the specific common construction reflects \*i, \*qi or both.

Table 2: Reflexes of \*i, \*qi and \*-ki in selected Oceanic languages

			Possessor morphemes				Classifier	Numeral	Free noun	Locative preposition
	specific personal		specific common		nonspecific (common)					
	inal	free	inal	free	inal	free				
POc	*k	*q	*i	–	*qi		*qi	*qi	*-ki	*i
Yapese	ʔk, ʔ	ʔ	–	–	ɛ: *i		–	i:, ɛ: *i	-kʰ   -y *ki   *i	–
Seimat	∅	∅	–	–	-i, -e		–	–	–	–
Kele	∅	k ∅∅	–	–	–		–	–	-y	–
Takia	∅∅k	∅∅k	–	–	–		–	–	-k *ki, *qi	–
Tigak	k, ∅	k, ∅	i	–	–	ina	–	–	–	–
Nochi	k, ∅	k, ∅	ina	ina	ina	ina/in <sup>1</sup>	–	ina	–	–
Lihir	k, ∅	k, ∅	–   -i	–   -i	–		–	i	–	i
Ramoaina	k, ∅	k, ∅	–	–	–	na	–	na <sup>2</sup>   in   ina	–	–
Tolai	k, ∅	k, ∅	-i	-i	–	na	–	na	–	–
Siar	k, ∅	k, ∅	–	–	in <sup>3</sup>		–	–	–	–
Halia	k, ∅	∅	–	–	–		–	–	–	i
Taiof	k, ∅	∅	–	–	in̄		–	–	–	–
Bogotu	k, ʔ	ʔ, ∅	–	–	i   – *qi, *i	–	–	–	–	i *qi, *i

<sup>1</sup> *in* only in construction with the small class of adjectives (see text).

<sup>2</sup> There are three noun classes in Ramoaina. Membership is lexically determined, although members of the *in* and *ina* classes tend to be noncount.

<sup>3</sup> Only in construction with adjectives (see text).

Table 2: (continued)

Kwaio	<i>k, ʔ</i>	<i>ʔ, ø</i>	–	–	–	<i>-ʔi, -ʔe</i> <i>*ki, *qi</i>	<i>-ʔe<sup>4</sup></i> <i>*ki, *qi</i>	–	<i>i</i> <i>*qi, *i</i>
Kwara'ae	<i>k, ʔ</i>	<i>ʔ, ø</i>	<i>i<sup>5</sup></i> <i>*qi, *i</i>	–	<i>ʔi</i>   – <i>*ki, *qi</i>	<i>-ʔi, -ʔe</i> <i>*ki, *qi</i>	–	–	<i>i</i> <i>*qi, *i</i>
Sa'a	<i>k, ʔ</i>	<i>ʔ, ø</i>	–	–	<i>i</i> <i>*qi, *i</i>	<i>i</i> <i>*qi, *i</i>	–	–	<i>i</i> <i>*qi, *i</i>
Arosi	<i>k, ʔ</i>	<i>ʔ, ø</i>	<i>i</i>   – <i>*qi, *i</i>	<i>i</i>   – <i>*qi, *i</i>	<i>i</i> <i>*qi, *i</i>	<i>i</i> <i>*qi, *i</i>	–	–	<i>i</i> <i>*qi, *i</i>
Mota	<i>ʔ, w, ø</i>	<i>ø</i>	<i>i<sup>6</sup></i>	–	<i>i</i>	–	–	<i>-i</i> <i>*ki<sup>7</sup></i>	<i>i</i>
Mwotlap	<i>ʔ, w, ø</i>	<i>ø</i>	<i>i<sup>8</sup></i>	–	–	–	<i>-ye<sup>9</sup></i> <i>*ki</i>	<i>-ye</i> <i>*ki</i>	–
Mwerlav	<i>ʔ, w, ø</i>	<i>ø</i>	<i>i<sup>10</sup></i>	–	–	–	–	<i>-[ʔ]i<sup>11</sup></i> <i>*ki</i>	<i>i</i>
N.E. Ambae	<i>k</i>	<i>ø</i>	<i>i, -i<sup>12</sup></i>	<i>-i</i>	<i>-i</i>	–	–	<i>-ki<sup>13</sup></i> <i>*ki</i>	–

<sup>4</sup> In *akwale-ʔe* 'ten of', from *akwala* 'ten' (Keesing 1985:88).

<sup>5</sup> Personal article used to indicate that the referent of the noun phrase is female.

<sup>6</sup> Article with personal names and nouns used as names.

<sup>7</sup> See text.

<sup>8</sup> Article with personal names and nouns used as names.

<sup>9</sup> See text.

<sup>10</sup> As Mota, but also in possessive, e.g. *na ak i Wok<sup>w</sup>as* 'W's canoe'.

<sup>11</sup> Contrast between *na-sasa-k* 'my name', *sesei* 'a name' (indep) and *na sese tadun* 'a man's name'.

<sup>12</sup> *i* is the personal article, *-i* what Hyslop calls the "construct suffix" attached to any possessed noun whose possessor is a noun phrase other than a pronominal.

<sup>13</sup> Codrington also records Volow (Saddle Island) *-ye*, Pak, Leon/Sasar, Mosin (all Vanua Lava) *-yi*.

Table 2: (continued)

Merei	∅	∅	i <sup>14</sup>	–	–	–	–	–	–
Tamambo	k, x	∅	–	-i *qi, *i	-i *qi, *i	–	-i <sup>15</sup> *qi, *i	–	–
Paamese	∅	∅	–	–	-i-	–	–	–	–
Nguna	k	∅	–	ki *ki	–	–	–	–	e- *qi, *i
Kwamera	k, ∅	∅	-i	–	i	–	–	–	i-
Anejoñ	ɣ	∅	-i *qi, *i	-i *qi, *i	–	–	–	–	–
Wayan Fijian	k	∅	i, -i *qi, *i	–	-i- *qi, *i	–	–	–	i *qi, *i
Bauan Fijian	k	∅	-i *qi, *i	–	-i- *qi, *i	–	–	–	e *qi, *i
Tongan	k	ʔ	–	–	–	-ŋ *qi	–	–	i / ŋ *i / *qi
E. Futunan	k	ʔ	–	–	ŋ *qi	ŋ *qi	–	–	i *i
E. Uvean	k	ʔ	–	–	ŋ *qi	ŋ *qi	–	–	ŋ *qi
Samoa	ʔ	∅	–	–	–	-i / -ŋ <sup>16</sup> *i, *qi / *ki	–	–	i *qi, *i

<sup>14</sup> Article with personal names and kin terms, also cliticised to a possessive classifier.

<sup>15</sup> Only with certain quantifiers, for example, *tua-i tamaloxi* 'some of the people', *tua-ra* 'some of them'.

<sup>16</sup> See text.

### 3.1 POc \*i/\*e ‘personal article’ outside possession phrases

Further support for the distinctness of \*i and \*qi is provided by the fact that each evidently served functions in Proto Oceanic outside possession phrases. Pawley (1972:32, 58) reconstructs \*i as a Proto Eastern Oceanic article occurring with personal nominals. He records reflexes with this function in certain Southeast Solomonic and North Vanuatu languages. (Some of these are noted in Table 2.) In Merei (North Vanuatu), for example, *i* always occurs as an article before a personal name (e.g. *i Pita* ‘Peter’) or a kin term (e.g. *i rabui-na* ‘his mother’). It similarly occurs as a personal article in Wayan Fijian (Andrew Pawley, pers. comm.). Pawley finds more widespread reflexes of \*i surviving only as a fossil prefix on independent pronouns and the word for ‘who?’ It seems very likely that Pawley’s Proto Eastern Oceanic \*i shares its origin with the personal article *e* that occurs in many Meso-Melanesian languages and is perhaps also reflected (John Lynch, pers. comm.) as a fossil prefix in certain Anejoñ (South Vanuatu) kin terms: *etpo-* ‘grandparent’ (POc \*tubu-), *etma-* ‘father’ (POc \*tama-), *etwa-* ‘same-sex sibling’ (POc \*tuqa- ‘older same-sex sibling’). Possibly \*e was the phrase-initial form, \*i phrase-internal.<sup>7</sup>

### 3.2 POc \*qi in numeral classifiers

Although Pawley (1972) does not mention \*qi as such, he does reconstruct two numeral classifiers, \*pua-qi/\*po-qi ‘spherical classifier’ and \*mata-qi ‘individual unit in series or class’. These forms are reflected in the Cristobal-Malaitan group of Southeast Solomonic, in Fijian and in Polynesian languages, and Hooper provides copious examples of them. In Cristobal-Malaitan there are far more classifiers, and they obviously originate in the inalienable nonspecific possession construction with \*qi. In Kwaio, for example, Keesing (1985:90) lists eleven classifiers, eight of which are recorded as nouns. They occur in phrases like (8), where *feʔe* (noun: *faʔi*) reflects POc \*puaq + qi (‘fruit’ + \*qi):

- (8) KWAIO  
*oru feʔe baʔu*  
 three fruit banana  
 ‘three bananas’

Similar phrases occur in Kwara’ae: *faʔi* in (9a) is cognate with Kwaio *feʔe*. But here the \*qi construction also remains productive, as in (9b). (Hooper’s examples are from Deck 1934.)

- (9) KWARA’AE  
 a. *faʔi niu*  
 NCL coconut  
 ‘a coconut’  
 b. *ʔae ʔi wae*  
 leg ʔi man  
 ‘human leg’

The reflexes of \*qi in classifiers are listed in the column labelled Classifier in Table 2.

<sup>7</sup> There is also evidence of another POc personal article \*qa, with reflexes in southern New Britain, Southeast Solomonic and Polynesian. How this contrasted with \*e/\*i is unknown.

Examples like (8) imply a Proto Oceanic numeral construction like the one in (10). The details are orthogonal to the present argument, but there is reason to think that POc numerals functioned both as adjectival verbs and as nouns. In their prenominal position they were predicates: hence the subject prefix \*i- in (10). Postnominally, they were attributes. Their nominal function is referred to below (Lynch, Ross & Crowley, forthcoming: ch. 4).

- (10) POC (?)  
 \*i-tolu puaq qi pudi  
 S:3S-three fruit \*qi banana  
 'three bananas'

This receives indirect support from constructions in certain other Oceanic languages.<sup>8</sup> There are a number of languages in which a reflex of \*qi intervenes directly between a numeral and a noun. Such reflexes of \*qi are listed in Table 2 under the heading Numeral. There are two possible sources of these constructions. First, in (11) the Kwaio numeral (*akwala* 'ten') is itself the head of an earlier nonspecific possession phrase ('two tens of things') in which it functioned as a noun:

- (11) KWAIO  
 rua akwale-e ʔola  
 two ten-ʔe thing<sup>9</sup>  
 'twenty things'

Second, in some languages an earlier classifier either is now fossilised or has disappeared entirely. Putative early Oceanic reconstructions are given beside the examples below in order to avoid lengthy discussion of issues that are not directly relevant here. In Yapese, the numeral *dalip* 'three' may reflect POc \*tolu 'three' + \*puaq 'general classifier' (see Ross 1996:149), with the structure of (10):

- (12) YAPESE POC (?)  
*dalip* ε: *ka:ro*: \*i-tolu puaq qi X  
 three ε: car S:3S-three NCL qi X  
 'three cars' 'three Xs'

Mwotlap (North Vanuatu) has two numeral constructions: (13a) is the normal numeral construction, whilst *yo-ye* in (13b) functions like a dual marker with animate nouns and is one of a set which also includes *etel-ye* 'trial' and *i-ye* 'plural' (Crowley, forthcoming). In the

<sup>8</sup> The presence of numeral classifiers in POc is attested by a scattering of languages with classifiers. As well as Cristobal-Malaitan and Polynesian, they include the Admiralties family, the Kilivila family, Sudest (Papuan Tip linkage), the North Bougainville linkage, the Nuclear Micronesian linkage and languages in New Caledonia. Reconstructable POc classifiers are \*puaq 'default classifier, round object' = 'fruit', \*kaiu 'wooden or elongated object' = 'tree' and \*tau 'person'. There are also numeral classifiers in non-Oceanic Austronesian languages in Indonesia. However, the variety of classifier constructions reflected in these languages suggests that the reconstruction here was only one of several early Oceanic constructions in which numerals and classifiers occurred. We find, for example, fossils like Tigak *potul* in a construction that reflects an early Oceanic ordering different from the one reconstructed in the text:

Tigak				Proto Oceanic
<i>ta</i>	<i>potul</i>	<i>a</i>	<i>nik</i>	*ta puaq tolu a/qi niuR
ART	three	LINKER	coconut	ART CL three ART/*qi coconut
	'three coconuts'			'three coconuts'

<sup>9</sup> *akwale-ʔe* loses its glottal stop when followed by a work beginning with a glottal stop (Keesing 1985:88).



second the classifier has apparently been lost from the construction in (10), an inference supported by the presence of a classifier in the first, albeit with a rearrangement of constituents:

- |      |                                |              |                                   |
|------|--------------------------------|--------------|-----------------------------------|
| (13) | MWOTLAP                        |              | POC (?)                           |
| a.   | <i>na-tm<sup>w</sup>an</i>     | <i>vo-yo</i> | *na tam <sup>w</sup> ane puaq rua |
|      | ART-man                        | vo-two       | ART man NCL two                   |
|      | 'two men'                      |              | 'two men'                         |
| b.   | <i>yo-ye tam<sup>w</sup>an</i> |              | *rua qi tam <sup>w</sup> ane      |
|      | two-ye man                     |              | two qi man                        |
|      | 'two men'                      |              | 'two men'                         |

Nochi and Tolai present us with a deletion of the classifier similar to (13b):

- |      |   |  |                       |
|------|---|--|-----------------------|
| (14) | NOCHI (Western Oceanic, Meso-Melanesian, New Ireland) <sup>10</sup> |  |                       |
|      | <i>sajaul a kuk ina niu</i>   |  |                       |
|      | ten ART one LIG coconut   |  |                       |
|      | 'eleven coconuts'   |  |                       |
| (15) | TOLAI   |  | POC                   |
|      | <i>a ivat na pal</i>  |  | *a pati qi pale       |
|      | ART four na house   |  | ART four *qi building |
|      | 'four houses'   |  | 'four buildings'      |

In Ross (1988:312–313) I showed that Maringe (Meso-Melanesian, New Ireland, Northwest Solomonic<sup>11</sup>) reflects the same structure in, as in *glimai khoilo*, where the final *-i* of *glimai* represents the *\*i-* of *\*ina* and the initial *kh-* of *khoilo* reflects the coalescence of the *-na* of *\*ina* with the initial *k-* of the citation form *koilo*.

### 3.3 Formal riddles in the reflexes of \*qi

Reflexes of *\*qi* in the modern languages are beset by two kinds of problem. The first concerns those reflexes in Table 2 which, on the face of it, do not look like reflexes of *\*qi* at all because they have the form *ina*, *in* or *na*. All of them occur in New Ireland languages. The second concerns the reflexes of the POC phoneme *\*q*.

I have discussed the first problem in Ross (1998b), and will only summarise that discussion here. The evidence suggests that in the language ancestral to the New Ireland linkage, the reflexes of *\*i* 'personal article' and *\*qi* fell together. Three events contributed to this. First, the personal article *\*i* was lost everywhere except in the possession constructions shown in the topmost cells of Table 1, and was thus reinterpreted as a possessive particle with a function similar to the function of *\*qi* with nonspecific (common) possessors. Second, the reflex of POC *\*qi* had extended its domain to include alienable possession, displacing *\*ni*<sup>12</sup> as it did in various other Oceanic languages. Third, POC *\*q* was lost, so that its reflex came to have the same form as the reflex of *\*i* 'personal article', in other words, both had the shape *\*i*. Thus putative early New Ireland structures were as follows:

<sup>10</sup> The ligature is used in Nochi only if the numeral is compound, i.e. greater than ten.

<sup>11</sup> The Northwest Solomonic group is part of the New Ireland linkage.

<sup>12</sup> POC *\*ni* survives almost nowhere in northwest Melanesia.

## (16) EARLY NEW IRELAND I:

	INALIENABLE	FREE
PERSONAL	*a ae=i X	*a Ruma na=i X
PRONOMINAL	*a ae-gu	*a na-gu Ruma
COMMON	*a ae-ñā tam <sup>w</sup> ata	*a na-ñā Ruma tam <sup>w</sup> ata
NONSPECIFIC	*a natu i boRok	*a polo i niuR

To the extent that this \*i indicated a case relationship, it resembled a preposition. Now, early Oceanic had at least one preposition which took a possessive suffix indicating the person and number of its governee. This was \*ta-, reconstructed by Pawley (1973) and Ross (1988:104–108), with some kind of locative, benefactive and possibly possessive function (Pawley 1973:148–149). Because of its usage, it was semantically close to early New Ireland \*i, and in some New Ireland languages \*ta- displaced the classifiers from free possessive constructions. However, reflexes of \*ta- formed a paradigm. In Tigak, for example, the form with a personal noun phrase is *te* (from \*ta- + \*i ‘personal article’); with a common noun phrase it is *tana* (from \*ta- + \*-ñā ‘P:3S’), and there are also forms reflecting \*ta-gu ‘1S’, \*ta-mu ‘2S’ and so on. Thus Tigak has a personal:common pattern *te:tana*, reflecting earlier \*tai:\*taña. The pattern is matched in Tigak by two other prepositions—*pe:pana* ‘instrumental’ and *su:suna* ‘allative’—and the pattern appears to be quite old. For \*i, however, the personal:common pattern was \*i:\*i, out of kilter with the other prepositions. This underwent analogical modification to \*i:\*ina. The fact that a cognate of Tigak *ina* occurs as far away as Taiof (north Bougainville), where *iñ* has exactly the same function, suggests that the pre-Tigak stage at which these things happened was in fact Proto New Ireland, the language ancestral to all languages of New Ireland and the northwest Solomons. The putative outcomes of these developments are depicted in (17):

## (17) EARLY NEW IRELAND II:

	INALIENABLE	FREE
PERSONAL	*a ae=i X	*a Ruma na=i X *a Ruma tai X
PRONOMINAL	*a ae-gu	*a na-gu Ruma *a Ruma ta-gu
COMMON	*a ae-ñā tam <sup>w</sup> ata	*a na-ñā Ruma tam <sup>w</sup> ata *a Ruma ta-ñā tam <sup>w</sup> ata
NONSPECIFIC		
EARLIER	*a natu i boRok	*a polo i niuR
LATER	*a natu ina boRok	*a polo ina niuR

The distribution of reflexes of \*i and \*ina among the categories of possessive construction outlined in §2 varies from one New Ireland language to another, as Table 2 shows. In Tigak, we find the distribution predicted by the account in the foregoing paragraphs: *i* with specific personal possessors, *ina* with nonspecific common ones. Not surprisingly, the ‘gap’ that this leaves in the system has been filled in some languages, and in Nochi, Ramoaina<sup>13</sup> and Tolai, for example, specific common possession is also marked by *i* or *ina*.

There is ample comparative evidence, presented in Ross (1998b), to show that Ramoaina and Tolai *na* reflect earlier New Ireland \*ina. Reflexes of \*ina have developed an important subfunction. Whilst their basic function is to link the possessed to the possessor in a

<sup>13</sup> The language of the Duke of York Islands.

nonspecific possessive construction, as in (18a), they also link an attribute to its noun, as in (18b). Although (18b) is the default 'adjectival' construction in Tolai, such evidence as the language provides us with indicates that *mamat* is the head of (18b). This means that (18a) and (18b) have the same structure.

## (18) TOLAI

- a. *a mapi na davai*  
 ART leaf LIG tree  
 'leaves of a tree'
- b. *a mamat na vat*  
 ART heavy LIG stone  
 'a heavy stone' (= 'a heavy one of a stone')

It also means that in Tolai and other South New Ireland languages *na* has an extended functional load, and has become paradigmatically divorced from other *i-* prepositional forms in the language. Because \**ina* arose by analogy, it was probably always monomorphemic, and the loss of initial *i-* is possibly a consequence of the high frequency of its use in the constructions in (18).<sup>14</sup>

If this history is correct, then we may, I think, claim *ina*, *in* and *na* as containing reflexes of POC \**qi*, even if by a rather complicated route.

The second formal riddle concerns the reflexes of POC \**q*. Table 2 presents (i) the reflexes of the putative Proto Oceanic morphemes with which this paper is concerned, (ii) the sound correspondences relevant to their phonological interpretation,<sup>15</sup> and (iii) the possible protoforms generated by applying the sound correspondences to each reflex. Reconstructions are shown as starred forms in cells beneath each morpheme. The possible Proto Oceanic forms from which the morphemes in the table are descended are \**i*, \**qi* and \**ki*. Where the sound correspondences are such that a morpheme could reflect any of these three forms, the morpheme is not a witness for phonological reconstruction, and no starred italicised forms are given. This means that by casting one's eye down a given column, it is easy to see phonologically contradictory reflexes.

Note that Table 2 does not include reflexes of morphemes other than those I am seeking to disambiguate here. Thus, although reflexes of \**qi* often alternate with those of \**ni* in daughter languages, I have omitted the latter here. Nor have I recorded zero reflexes, as they tell us nothing useful.

We can see from Table 2 that the reflexes of POC \**i* 'personal article' do not allow us to distinguish between \**qi* and \**i*. That is, the Proto Oceanic form may, on its Oceanic reflexes, have been \**qi*. The reflexes of the possible alternant form \**e* mentioned above also do not

<sup>14</sup> The irregular deletion of a segment in a morpheme with high token frequency is not surprising. It is comparable to the irregular deletion of Latin *-t-* in the second person plural inflection of Spanish verbs (Latin *-atis* > *-ades* > *-ades* > *-áis*; Bybee 1994).

<sup>15</sup> Sources of sound correspondences are, for Yapese, Ross (1996); for languages from Seimat to Halia, Ross (1988); for languages from Bugotu to Arosi, my own analysis; for languages from Mota to Nguna, Tryon (1976) as well as Jauncey (1997) for Tamambo; for Kwamera and Anejoñ, John Lynch (pers. comm.) and for Fijian and the Polynesian languages Geraghty (1986), as well as for Polynesian languages, Clark (1973). My interpretation of the reflexes of \**q* in Cristobal-Malaitan languages (Kwaio, Kwara'ae, Sa'a, Arosi) differs somewhat from that of Lichtenberk (1988). Where he infers that POC \**q* is always lost in these languages, I infer that it was occasionally retained throughout Southeast Solomonian, for example initially in reflexes of POC \**quwe* 'rattan' and \**qiri(s)* 'cut up' and medially in reflexes of \**qaqe* 'leg'.

allow us to distinguish between \*qe and \*e. The only reflex which suggests something different is Nguna (Central Vanuatu) *ki*, reflecting earlier \*ki. However, as this occurs only with specific common possessors (Schütz 1969:41–42), it is an unlikely reflex of POc \*i. Moreover, it is clear from Ray (1926:217–218) that this is not a reflex of \*qi but is cognate with the forms from which Pawley (1972:85) reconstructs the dative Proto Eastern Oceanic preposition \*ki. I return to the form of POc \*i below.

I assume on functional grounds that the columns labelled nonspecific, Classifier and Numeral in Table 2 all reflect the same POc morpheme \*qi, as argued in §3.2. Examining the overall pattern of reflexes of \*qi, we see that a majority of the criterial reflexes reflect \*qi or \*i, that Kwaio and Kwararʼæ forms reflect \*ki or \*qi,<sup>16</sup> and that the Tongan, East Futunan and East Uvean (all Polynesian) forms reflect \*qi.<sup>17</sup> This indicates that the Proto Oceanic form was indeed \*qi.

There are three contrary voices. The first two have limited significance. They are Yapese, apparently reflecting \*i, and the fossilised reflex in Mwothlap number markers (see (13) above). Discussion of Yapese is postponed until later (§6), and the Mwothlap reflex is discussed in association with the reflexes of the free noun morpheme in §4.

The third contrary voice is in Samoan, where there are two fossil reflexes—(expected) *-i* and (unpredicted) *-ā*, reflecting \*qi and \*ki, respectively. Neither is productive. Both occur in lexicalised compounds containing reflexes of Proto Polynesian \*fua-qi- (from POc \*puaq qi) and Proto Polynesian \*mata-qi- (not reconstructable for Proto Oceanic). The suffix *-i* is found in, for example, *fuaiʻāpu* ‘sentence’ (*ʻāpu* ‘word’), *fuaitau* ‘words or lines of a song’ (*tau* ‘count’ VERB), *fuaiala* ‘part of a village’ (*ala* ‘path’), *mataitōʻeta* ‘most valuable fine mat (in a collection)’ (*tōʻeta* ‘fine mat’), *mataitayata* ‘fine-looking man’ (*tayata* ‘man’). The suffix *-ā* is found in such forms as *fuaʻāfaʻā* ‘a single banana’ (*faʻā* ‘banana’), *fuaʻāvai* ‘a single water bottle’ (*vai* ‘coconut water bottle’), *mataʻāʻololoa* ‘the best article among a lot of goods’ (*ʻololoa* ‘goods’), *mataʻāsiva* ‘best dancer in a night dance’ (*siva* ‘dance’ VERB) (Mosel & Hovdhaugen 1992:242). Mosel and Hovdhaugen gloss *fuai-* and *fuaʻā-* differently, the former as ‘a collection or group of identical objects’, the latter as ‘a single piece of a kind’, but this does not assist us in reconstructing the history. They gloss both *matai-* and *mataʻā-* as ‘typical or prominent representative of something’.

The most obvious explanation for the unpredicted form *-ā* is that it is a variant of *-i* with glottal stop epenthesis. This may seem unmotivated, but there is evidence elsewhere in Samoan of glottal stop epenthesis: the dual pronouns *tāʻāua* ‘D:1D’, *māʻāua* ‘D:1ED’ and *lāʻāua* ‘D:3D’ are reflexes of Proto Nuclear Polynesian \*tāua, \*māua and \*lāua.<sup>18</sup> If epenthesis occurred here, then it may have occurred in the forms *fuaʻā-* and *mataʻā-*.

Since, as Hooper (1985) has shown, POc \*i ‘personal article’ and \*qi ‘nonspecific possessive particle’ contrasted within the possessive system in POc, we have circumstantial—

<sup>16</sup> There are Kwaio noun phrases like *taʻā i asi* ‘people of the sea’ and *aliola i ʻAleʻale* ‘Are’are canoes’ which appear superficially to reflect the POc nonspecific possessive construction with \*qi, but, as Keesing (1985:100) points out, the attributes in these phrases are always locative, and *i* is the locative preposition, not a possessive marker.

<sup>17</sup> I have decided to ignore Polynesian forms that reflect the Proto Polynesian classifiers \*fuaa and \*mataa (see Hooper 1985), on the grounds that these appear to be ancient alternants of \*fua-qi and \*mata-qi but were not derived from them and did not reflect POc \*qi.

<sup>18</sup> I am grateful to Andrew Pawley for drawing my attention to this.

but only circumstantial—evidence that, since the latter was clearly \*qi, the former was probably \*i, despite the ambiguity of its reflexes.

#### 4 Distinguishing between \*qi ‘nonspecific possessive particle’ and \*-ki ‘free noun suffix’

Hooper attributes three functions to \*qi, two of which are discussed thoroughly. The first is the possessive function described above in §2 and §3. The second is the classifier function discussed in §3.2. Hooper also refers briefly on two occasions to the suffix that in some northern Vanuatu languages forms a ‘free’ version of inalienable nouns. She writes (1985:156):

It is reasonable to assume that in North Vanuatu languages an inalienable possessive particle \*qi ‘became attached to the noun as a suffix and was then reinterpreted, either as part of the noun base, or as a gender marker rather than a possessive’ (Pawley 1972:115).

I would like to suggest that in the light of now available data this assumption is not quite as reasonable as it seems. Semantically, it is perhaps plausible, but the phonological form of the reflexes in the Free noun column of Table 2 suggests that in fact Proto Oceanic had two morphemes, possessive \*qi and free-form derivative \*-ki<sup>19</sup>.

The relevant northern Vanuatu languages have a free-noun suffix with the forms Mwotlap -ye, Mwerlav, North-East Ambae -ki, unambiguously reflecting POc \*-ki. The most crucial evidence comes from North-East Ambae (and has only recently become available), where there is a contrast between -i, marking possession with all three kinds of possessor, and -ki, the free-form derivative suffix that occupies the same slot as the pronominal possessor suffixes on inalienable nouns, giving contrasts like *vulu-gu* ‘my hair’ vs *vulu-ki* ‘hair (unpossessed)’. The contrast between -i and -ki is equally clear:

##### (19) NORTH-EAST AMBAE

- a. *Mo toka lo ulu-mu.*  
 REALIS sit PREP above-P:2SG  
 ‘She is sitting on top of you.’
- b. *Mo toka lo ulu-i gai.*  
 REALIS sit PREP above-i tree  
 ‘She is sitting on top of a tree.’
- c. *Mo toka lo ulu-ki.*  
 REALIS sit PREP above-FREE  
 ‘She is sitting on top.’

- (20) ... *mo tuli na vinu-ki ta lolo tahi.*  
 REALIS throw ART skin-FREE PREP inside sea  
 ‘...[Suqe] was throwing the skins into the sea.’

<sup>19</sup> Pawley (1972:115) does reconstruct \*ki, but as a Proto Eastern Oceanic possessive particle form that occurred “after nouns of ... edible gender ... when possessed by a personal name and possibly in certain other contexts”. I do not have appropriate evidence to reconstruct this particle in POc, and it would in any case be irrelevant to the present discussion, where inalienable nouns are under consideration.

*Ale Takaro mo lehe na vinu-i ka-na mena ...*  
 so Takaro REALIS see ART skin-i CL-P:3S ripe  
 'So Tagaro saw the skins of his ripe bananas ...'

The North Vanuatu language for which Codrington (1885) provides the most complete data is Mota. This happens to be the language where the free-form suffix is *-i*, in other words, it is apparently not distinguishable from the reflex of *\*qi*. However, a careful reading of Codrington (1885:261–262) suggests that they are distinct. There is a morphophonemic difference between *-i* 'nonspecific possessive' and *-i* 'free form suffix', as in *sase tanun* 'man's name' (where *sase* represents *sasa-* + *i* 'possessive') vs *sasa-i* 'name (free form)'. One may infer that *sase* reflects a development whereby POC *\*-a-qi* became *\*-ai*, then *\*-e*, whereas *sasai* reflects a later set of changes such that POC *\*-a-ki* became *\*-aʔi* and then *-ai*.

The reflexes of *\*-ki* 'free-form suffix' mentioned above are all from North Vanuatu languages, and these do not justify a Proto Oceanic reconstruction. There are two other reflexes of *\*-ki*, however, as well as a possible third. The two are in Takia (Western Oceanic, North New Guinea linkage) and in Kele (Admiralties). In Takia we find, for example, *nanu-g* 'my child', *nanu-n* 'her/his child', but *nanu-k* 'a child' (possessor unidentified); and in Kele *leme-m* 'your arm', but *leme-y* 'arm (free form)'. Unfortunately, the Takia and Kele forms provide no phonological disambiguation, as POC *\*k* and *\*q* both become Takia *-k* when they are reflected word-finally after the loss of a Proto Oceanic final vowel, and they are both deleted between vowels in Kele. However, their function indicates that they reflect the same morpheme as the North Vanuatu reflexes.

The other possible reflex of *\*-ki* is in Yapese, where Jensen (1977:143) reports *ʔaʔi:-y* 'my liver', *ʔaʔi:-m* 'your (s) liver', *ʔaʔi:-n* 'her/his liver', *ʔaʔi:-y* 'anyone's liver'. The suffix *-y* on the last item of the series is a free-form suffix, but it is not clear if it is a phonologically plausible reflex of POC *\*-ki* (see §6).

Free-form suffixes are also reported by Hollyman (1991) and Ozanne-Rivierre (1991) in the languages of north New Caledonia. However, the consonants of the ones that look as if they might be reflexes of *\*-ki*, namely Kumak *-t*, Nyâlâyu (Belep) *-t* and Nyâlâyu (Balade) *-r*, reflect POC *\*s* or *\*c*, not *\*k* (for which the regular reflexes are Kumak *-c*, Nyâlâyu *-∅* (Hollyman 1991:148–150)).

I noted above that Mwotlap *-ye*, an apparently fossilised reflex of *\*qi* attached to certain numerals, is phonologically irregular, reflecting POC *\*ki*. It appears that, since *\*qi* has no productive reflexes in Mwotlap, the reflex of *\*qi* has merged irregularly with that of *\*-ki*.

## 5 POC *\*i* 'locative preposition'

The rightmost column of Table 2 contains reflexes of POC locative preposition *\*i*, perhaps the only true preposition that occurred in Proto Oceanic (Ross 1988:104). All reconstructions to date have the form *\*(q)i* (Pawley 1972:85, 1973; Ross 1988:104), raising the question of whether this morpheme had an initial consonant. The evidence now available suggests that it did not. Moysse-Faurie (1993:178) reports a contrast in East Futunan between *i* 'locative preposition' and *ʔ* 'possessive relator'. The latter, in examples such as *kili ʔ manu* 'animal skin', is clearly a reflex of the nonspecific possessor function of POC *\*qi*. On this evidence, POC locative preposition *\*i* may be reconstructed in contrast with possessive *\*qi*.

The only contrary evidence occurs in Tongan, where both *i* and *ĩ* occur as reflexes of the locative preposition. Clark (1973:22–23) accounts for *ĩ* as the outcome of glottal stop prothesis in phrase-initial position.

## 6 A note on Yapese reflexes

I tabulate putative Yapese reflexes of POc \*qi and \*-ki in Table 2 simply because it is noteworthy that Yapese seems to reflect them. Two possible reflexes of the free-form suffix \*-ki are noted. One of these, *-y*, was exemplified in §4 and is functionally a free-form suffix but is not an expected reflex of \*-ki. The other, *-k*, is a phonologically adequate reflex,<sup>20</sup> but does not serve as a free-form suffix. Instead it replaces *-n* ‘P:3S’ on just two morphemes, the prepositions *ro:-* ‘possessive’<sup>21</sup> (*ro:-γ* ‘my’, *ro:-m* ‘your (S)’, *ro:-k<sup>2</sup>* ‘her/his’, *riy* ‘its’) and *ŋo:-* ‘benefactive, directional’ (*ŋo:-γ* ‘to me’, *ŋo:-m* ‘to you (S)’, *ŋo:-k<sup>2</sup>* ‘to her/him’, *ŋa:y* ‘to it’). However, the *-k* forms are clearly third person singular in function, not free forms. Both prepositions have a ‘free form’ in *-y*, used when the governed noun phrase is not mentioned (Jensen 1977:149–150, 190).

Perhaps as many as five layers of Austronesian lexical items can be identified in Yapese. The Yapese reflexes of POc \*k and \*q given in Table 2 are from what I call Set C (Ross 1996), the Oceanic set reflected in most Yapese grammatical morphemes (and a considerable part of the lexicon) and thought to be the earliest layer of the language. However, the putative Yapese reflexes of POc \*qi and \*-ki do not match the relevant sound correspondences. There are two possible reasons for this. First, historical research into Yapese is at an initial stage, and the analysis in Ross (1996) may be inadequate. Second, Yapese may on its own constitute a first-order subgroup of Oceanic. If so, it is possible, for example, that \*qi did not acquire its \*q- until after Yapese had separated from the rest of Oceanic. But this is speculation.

## 7 Conclusion

I conclude that the morphemes presented by Hooper (1985) and others cognate with them reflect four Proto Oceanic morphemes:

*i	personal article
*qi	nonspecific inalienable possessive marker
*-ki	free-form derivative suffix
*i	locative preposition

It is more difficult than one might expect to find non-Oceanic cognates of these morphemes in languages where initial Proto Austronesian \*q- is reflected unambiguously.

Only Proto Austronesian \*i ‘locative preposition’ can be reconstructed with certainty, as there are ample reflexes of it, and these include reflexes in the Formosan languages Thao, Amis and Paiwan, where \*qi would be reflected differently from \*i (see Blust 1995).

<sup>20</sup> *k<sup>2</sup>* does not occur among the Yapese reflexes recorded in Ross (1996) but glottalisation in Yapese is frequently the result of conditioning by an adjacent rounded vowel.

<sup>21</sup> Possibly a reflex of POc preposition \*ta-, but initial *r-* reflects POc \*r-, not \*t-.



Proto Austronesian \*i 'personal article' is reconstructable with somewhat less certainty, because its would-be reflexes vary somewhat in function and it is unsafe to reconstruct a morpheme of such small phonological substance without functional correspondence.

There are no known potential cognates of \*qi in Formosa, so if \*qi is reconstructable at an interstage earlier than Proto Oceanic, it would not be Proto Austronesian. There are a number of plausible cognates, but all occur in Western Malayo–Polynesian languages where initial \*q- is lost, and a majority co-occur only with personal possessors, suggesting that they are reflexes of \*i 'personal article', which have undergone the same functional limitation to possessive phrases as we observed in various Oceanic languages above. Blust (1977, 1995) notes this problem.

As yet, I have found no non-Oceanic cognate of \*-ki 'free-form derivative suffix'.

Apart from confirming Hooper's findings with regard to \*i 'personal article' and \*qi, the new contributions made by this paper are to reconstruct POC \*-ki 'free-form derivative suffix' and to find that the Proto Oceanic locative preposition was \*i rather than \*qi, and therefore not the same morpheme as possessive \*qi.

## References

- Blust, Robert A., 1977, The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics* 9(2):1–15.
- 1995, Austronesian comparative dictionary. Computer files. University of Hawai'i, Honolulu.
- Bybee, Joan L., 1994, A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics* 5:285–305.
- Clark, Ross, 1973, *Aspects of Proto-Polynesian syntax*. Auckland: Linguistic Society of New Zealand.
- Codrington, Robert H., 1885, *The Melanesian languages*. Oxford: Oxford University Press.
- Crowley, Terry, forthcoming, Mwotlap. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*. Richmond: Curzon Press.
- Deck, Norman C., 1934, *A grammar of the language spoken by the Kwara'ae people of Mala, British Solomon Islands*. Reprint No. 5 from JPS 42, 43. New Plymouth: Thomas Avery.
- Dixon, R.M.W., 1988, *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- Geraghty, Paul, 1986, The sound system of Proto-Central-Pacific. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds, *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 289–312. Canberra: Pacific Linguistics.
- Hollyman, K.J., 1991, Personalised and non-personalised possession: final consonants in Kumak and other languages of Far Northern New Caledonia. In Robert A. Blust, ed. *Currents in Pacific Linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 145–154. Canberra: Pacific Linguistics.
- Hooper, Robin, 1985, Proto-Oceanic /*\*qi*/. In Andrew K. Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 141–167. Canberra: Pacific Linguistics.



- Jauncey, Dorothy, 1997, A grammar of Tamambo, the language of western Malo, Vanuatu. PhD dissertation, Australian National University.
- Jensen, John Thayer, 1977, *Yapese reference grammar*. Honolulu: University Press of Hawai'i.
- Keesing, Roger M., 1985, *Kwaio grammar*. Canberra: Pacific Linguistics.
- Lichtenberk, Frantisek, 1985, Possessive constructions in Oceanic languages and Proto-Oceanic. In Andrew K. Pawley and Lois Carrington, eds *Austronesian linguistics at the 15<sup>th</sup> Pacific Science Congress*, 92–140. Canberra: Pacific Linguistics.
- 1988, The Cristobal–Malaitan subgroup of Southeast Solomonic. *Oceanic Linguistics* 27:24–62.
- Lynch, John, 1982, Towards a theory of the origin of the Oceanic possessive constructions. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics 1: Currents in Oceanic*, 243–268. Canberra: Pacific Linguistics.
- 1996a, “Passive”, “characteristic” and “food” possession in Oceanic languages. Unpublished ms. University of the South Pacific, Port Vila.
- 1996b, Proto Oceanic possessive marking. In John Lynch and Fa'afu Pat, eds *Oceanic studies: proceedings of the First International Conference on Oceanic Linguistics*, 95–112. Canberra: Pacific Linguistics.
- Lynch, John, Malcolm Ross and Terry Crowley, eds, forthcoming, *The Oceanic languages*. London: Curzon Press.
- Mosel, Ulrike and Even Hovdhaugen, 1992, *Samoan reference grammar*. Oslo: Scandinavian University Press.
- Moyse-Faurie, Claire, 1993, *Dictionnaire futunien-français avec index français–futunien. Langues et Cultures du Pacifique* 8. Paris: Peeters.
- Ozanne-Rivierre, Françoise, 1991, Incorporation of genitive relators in the languages of New Caledonia and the Loyalty Islands. In Robert A. Blust, ed. *Currents in Pacific Linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 321–338. Canberra: Pacific Linguistics.
- Pawley, Andrew K., 1972, On the internal relationships of eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history* 3, 1–142. Honolulu: Bernice Pauahi Bishop Museum.
- 1973, Some problems in Proto Oceanic grammar. *Oceanic Linguistics* 12:103–188.
- Pawley, Andrew K. and Timoci Sayaba, 1990, Possessive-marking in Wayan, a western Fijian language: noun class or relational system? In J.H.C.S. Davidson, ed. *Pacific island languages: essays in honour of G.B. Milner*, 147–171. London and Honolulu: School of Oriental and African Studies and University of Hawai'i Press.
- Ray, Sidney H., 1926, *A comparative study of the Melanesian island languages*. Cambridge: Cambridge University Press.
- Ross, Malcolm D., 1988, *Proto Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.

- 1996, Is Yapese Oceanic? In Bernd Nothofer, ed. *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, 121–166. Hamburg: Abera.
- 1998a, Proto-Oceanic adjectival categories and their morphosyntax. *Oceanic Linguistics* 37:85–119.
- 1998b, Possessive-like attribute constructions in the Oceanic languages of northwest Melanesia. *Oceanic Linguistics* 37:234–276.
- Schütz, Albert J., 1969, *Nguna grammar*. Oceanic Linguistics Special Publication No. 5. Honolulu: University of Hawai'i Press.
- Tryon, D.T., 1976, *New Hebrides languages: an internal classification*. Canberra: Pacific Linguistics.

# *Semantic and syntactic functions of reduplication in Niuean*

---

WOLFGANG B. SPERLICH

An analysis of the semantic and syntactic roles of reduplication in Niuean is presented here. As only a few detailed descriptions of reduplication for Polynesian languages exist, especially from a semantic and syntactic perspective, the present study relies on Niuean data alone. The main categories established are those of semantic and syntactic triggers with intermediate stages. Notions of frequency, repetition and plurality serve as main descriptors. Verb–argument agreement is the main syntactic function. Noun plurals account for only a small class. The study also addresses questions of productivity and what the possible lexical selection criteria for reduplication could be. Finally the difficulties of establishing the base of a reduplication—be it synchronic or diachronic—are exemplified. A brief conclusion notes the need for further analysis if detailed descriptions are to be linked to derivational rules.

## 1 Introduction<sup>1</sup>

In general, linguistics in-depth descriptions of the phenomenon of reduplication have mainly focused on phonological/morphological levels (Marantz 1982, McCarthy and Prince 1995) and the same is true for Polynesian languages where reduplication is well attested (see Meyerhoff & Reynolds 1996 for Maori). Equally, the very substantial grammar of Samoan (Mosel & Hovdhaugen 1992) treats reduplication under ‘morphology’, noting in passing what the syntactic/semantic functions are. In contrast, Elbert and Pukui’s (1979) *Hawaiian Grammar* gives over a chapter to the ‘meaning of reduplications’ separate from the phonological/morphological treatment. In this work, I will attempt to give priority to the question of what syntactic/semantic environments trigger reduplication—for Niuean. Of course, there is an important interface between syntactic/semantic environments and the phonology/morphology of items potentially undergoing reduplication, in that the latter will

---

<sup>1</sup> It is indeed both an honour and a pleasure to contribute to Byron Bender’s Focusschrift, especially as I had the good fortune to publish the new Niuean dictionary under his general editorship of the PALI Language Texts. Byron has not only encouraged me in my work, but has done so countless times to many others, and together with his own work has contributed to Oceanic linguistics like no one before him. *Fakaauē lahi!*

resist or aid in the formation of reduplicated forms in varying degrees, but here we will restrict the discussion to environments that actually have triggered processes of reduplication. Still, within this inquiry there is the problem of being able to ascertain if a reduplication process has actually been involved, especially if the base form is not extant but is transparent in related derivational forms. In such cases phonological/morphological criteria may play a role.

Previous studies of Niuean have noted the importance of reduplication but have not analysed the phenomenon in depth. McEwen (1970:xi) lists the functions as “to form a plural, to convey the repetition of an action, to lessen the force, or in some other way to modify the basic meaning of the word”. Seiter (1980:62) notes that “many other intransitive verbs have plurals formed by reduplicating the first syllable of their singular form” and “a few transitive verbs in Niuean obligatorily agree in number with their direct object. For some, the plural is formed by reduplicating the first syllable of their singular form”. Massam and Roberge (1997) remark on the possibility that number agreement is not really a kind of grammatical agreement but rather an aspectual type.

## 2 A syntactic/semantic trigger for generating reduplicated forms in Niuean

The notion of ‘trigger’ as used here and elsewhere in this paper refers to syntactic and/or semantic determinants (see Bybee 1985, ch. 2, who talks of “semantic determinants of inflectional expressions”; and Lieber 1992:179, who uses the word “trigger” in an example where an application of one rule ‘triggers’ the application of another). I do not want to claim here that ‘meaning’ determines ‘form’, but merely point out what the possible relationships are with regard to reduplication in Niuean. The proposed processes are based on the premise that all phonological and morphophonemic conditions for reduplication have been met. In line with many other Oceanic languages, I will use the label ‘REdup’ as shorthand for reduplication of the first mora of the base, and ‘reDUP’ for reduplication of the last two morae. The label DUP, for full reduplication, can be considered an instance of reDUP where the base consists of only two morae (Rehg, pers. comm.).

$$(1) \quad ( \text{---} ) \text{ LEXICAL ROOT } [x,y] ( \text{---} ) \rightarrow ( \text{---} ) \{ \text{reDUP} \} [x,y] ( \text{---} )$$

DUP  
REdup

where (---) = optional affixation/derivation  
 [x,y] = syntactic/semantic categories  
 → = diachronic/synchronic derivation

A typical example is provided in (2), where both full and partial reduplication (REdup and reDUP are exclusive of each other) occur.<sup>2</sup>

- (2)     *aku*, v.t. to dig up, to scoop ... [PPN \*aku ‘scrape out with hands’]  
           *aaku*, REdup, v.t. to scoop up once ...  
           *akuaku*, DUP, v.t. to dig, scoop repeatedly ...

<sup>2</sup> All data from Sperlich (1997 and fieldnotes, n.d.). Abbreviations used herein include: n. – noun, PPN – Proto Polynesian, Prf. – prefix, Suf. – suffix, v.i. – verb intransitive, v.t. – verb transitive.

With verbs that imply movement performing some action, the progression from REdup (or reDUP) to DUP often mirrors 'once' to 'repeatedly', while the unreduplicated base form is nonspecific. In other cases, however, no such semantic functions can be established, as in (3), where *aagi* and *agiagi* are synonyms.

- (3) *agi*, v.i. 1. to blow (of wind) ... 2. to wander, to be a vagabond ...  
     n. nonplaying member of a cricket team ...  
     [PPN \*agi 'blow (of wind, breeze)'; \*agi 'unencumbered, unhampered']  
*aagi*, REdup, v.i. to blow gently (of wind) ...  
*agiagi*, DUP, v.i. to blow gently (of wind) ...

### 3 Synchronic reduplication processes

If we make it a strict requirement that any reduplicated form **must** be derived from root forms that occur unreduplicated on the surface, then we can exemplify the following permutations.<sup>3</sup>

#### 3.1 Semantically (and syntactically) conditioned

- (4) *apo*, v.i. to beg ...  
*apoapo*, DUP, v.i., FREQUENTIVE, to keep begging ...
- (5) *tagina*, v.i. to be dazed ...  
*taginagina*, reDUP, v.i., INTENSIVE, to be really dazed ...  
 \*gina, \*ginagina
- (6) *aafu*, v.i. to be sultry, to be hot ... [PPN \*qafu 'to be hot and humid']  
*afuafu*, (?)DUP, 1. v.i. to be hot ... SEMANTIC FIELD RESTRICTION  
 2. v.t. to dampen ... SEMANTIC CHANGE / ?SYNTACTIC CATEGORY  
 CHANGE [PPN \*afu-afu 'drizzle, light rain' (*uha*, n. rain ... PPN \*quha 'rain')]  
 \*afu
- (7) *ano*, n. a trace of something ... [PPN \*ano 'desolate']  
*anoano*, DUP, n. trace of twilight .... SEMANTIC RESTRICTION
- (8) *atu*, collective particle, row, array, group ... [PPN \*qatu 'line, row']  
*fakaatu* (*faka-atu*), v.t. to put in a row ...  
*fakaatuatu*, Prf., (?)DUP/reDUP, v.t., FREQUENTIVE, to put in many rows ...  
 (?)SYNTACTIC CATEGORY CHANGE  
 \*atuatu
- (9) *ana*, n. cave, den ... [PPN \*qana 'cave']  
*taanaana* (*ta-ana-ana*), Prf., DUP, v.i. to be hollow, bare, open ... SEMANTIC  
 CHANGE / SYNTACTIC CATEGORY CHANGE  
 \*anaana, \*taana

<sup>3</sup> The syntactic/semantic functions are given in capitals; starred items outside the square brackets indicate that the word does not occur in this form, even though it could be expected to occur; starred items inside the square brackets refer to protoforms.

The semantic triggers FREQUENTIVE and INTENSIVE are common to reduplication in many Polynesian languages (see Krupa 1982). However, the instances of SEMANTIC RESTRICTION/CHANGE are interesting because the semantic field of the base is thereby narrowed down to a more specific meaning or the semantic field is selectively extended in some way. These processes can be called extensions of the lexicon itself—as a further strategy to overcome the limited morpheme base due to the small phonemic inventory (see Krupa 1982, ch. 3). Of interest here are also the cases where a semantic change is accompanied by a syntactic category change. The latter seems to occur mainly when the base word is a category other than a verb. (The occurrence of a valency change, as in (8), may only be termed an internal category change if it is any change at all.)

Semantically conditioned reduplication as 'lexical extension' would have few formal rules, just as innovation in the base lexicon is not generally rule driven.

### 3.2 Syntactically (and semantically) conditioned

Under this category, we expect much more rule-governed behaviour, at least as far as the syntactic trigger is concerned. The question of how lexical items are selected (and others not) remains largely unanswered. The essential processes are: (i) reduplicating verbs to show agreement with plural subjects or objects, (ii) reduplicating nouns to change singulars into plurals, and (iii) reduplicating verb stems to derive nouns. The latter two occur fairly rarely as alternatives to other common syntactic devices, such as the use of *tau* as a preposed indicator of plural nouns.

Example (10) illustrates the first process. The verb is intransitive, so the notion of 'subject' may be redundant (but see the transitive verb below). Note also that we can be confident that *takoto* is a verb stem (not derived from *ta-koto*) since the first syllable is reduplicated. (Derived stems are not subject to reduplication in Niuean.)

- (10) *takoto*, v.i. to lie down ... [PPN \**takoto* 'lie down']  
*Kua takoto a ia ke mohe.* 'He lay down to sleep.'  
*tatakoto*, REdup, v.i. to lie down ... SUBJECT PLURAL  
*Kua tatakoto a laua ke mohe.* 'They lay down to sleep.'

Example (11) shows the use of DUP and REdup simultaneously indicating syntactic change and semantic field restriction, perhaps best described as a portmanteau phenomenon.

- (11) *tali*, v.t. to meet, greet, welcome, wait for ... [PPN \**tali* 'wait']  
*talitali*, DUP, v.t. to expect ... SEMANTIC RESTRICTION  
*tatali*, REdup, v.t. to wait for ... SEMANTIC RESTRICTION / SUBJECT PLURAL  
*Kua tatali a lautolu ke taā e logo.* 'They waited for the bell to ring.'  
*fakatali*, v.t. to wait (to be made to wait), to expect ...  
*fakatalitali* (*faka-tali-tali*), Prf., DUP, v.t. to wait (to be made to wait), to expect ...  
 SUBJECT PLURAL  
*Kua fakatalitali a lautolu he matua ke hau he vao.* 'They waited for their father to return from the bush.'  
 \*fakatatali

- (12) *taaki*, v.t. to uproot, to pull out/up ... [PPN \**taqaki* 'pull up or out, hoist, extract']  
*Kua taaki e au e huli talo.* 'I pulled out a taro shoot.'  
*tataaki*, REDup, v.t. to uproot, to pull up/out ... OBJECT PLURAL  
*Kua tataaki e au e tau huli talo.* 'I pulled out the taro shoots.'  
 \**taakitaaki*

In (13), both SUBJECT PLURAL and OBJECT PLURAL (as well as other features) occur in the same derivational paradigm.

- (13) *maga*, v.i. to be forked, branched, parted (in two parts) ...  
 n. 1. fork (in tree) ... 2. part, division, piece ...  
 [PPN \**maga* 'branch, fork; branching, forked']  
*magamaga*, DUP, v.i. 1. to mature ... 2. to broaden ... 3. to be unfinished ...  
 n. 1. fork (of tree) ... 2. crotch ...  
*mamaga*, REDup, v.i. to gradually divide ...  
*fakamaga*, v.t. 1. to open one's mouth ...  
 2. to make a fork (for hooking down fruit) ...  
*fakamagamaga*, v.t. to open one's mouth ... SUBJECT PLURAL  
*fakamamaga*, v.t. 1. to spread out, to open up or out ... 2. to gape ...  
*magai (maga-i)*, Suf., v.t. to place between, to interpose ...  
*magamagai (maga-maga-i)*, DUP, Suf., v.t. to place between, to interpose ...  
 OBJECT PLURAL  
*Kua magamagai e ia e tau koloa tui haana he tau matahio.* 'He placed his clothes between the louvre windows.'

The next example involves a category change from verb to noun, as well as a semantic change.

- (14) *ako*, v.i. to learn ... [PPN \**ako* 'learn, teach']  
*akoako*, DUP, v.i. to learn steadily ...  
 n. pastor ...

The next couple of examples demonstrate the (quite rare) occurrence of syntactic change (number) within nouns.

- (15) *tepu*, n. lump, knot, wart, knob, clitoris ...  
*teputepu*, DUP, n. lumps, knots, warts, knobs ... PLURAL
- (16) *alo*, n. (local noun), under, inside (of a surface) ... [PPN \**qalo* 'belly, bowels']  
*aloalo*, DUP, n. (local noun), under, inside (of surfaces) ... PLURAL

There are a few interesting cases where verbs have plural suppletives. If either of the verb forms is reduplicated (or has any other morphology), it is still by definition either singular or plural; for example:

- (17) *kata*, v.i. to laugh ... subject singular, plural suppletive is *feki*: ...  
*katakata*, DUP, v.i. to be happy, to smile ... SUBJECT SINGULAR

Krupa (1982) notes that partial reduplication in Polynesian languages is generally indicative of verb–nominal argument agreement demonstrated here. In Niuean, there is certainly a trend in that direction, but not exclusively in that direction. Seiter's (1980) possible implications that subject plural triggers only intransitive verbs is not supported by

my data. That object plural triggers only transitive verbs is clear by definition. Massam and Roberge (1997) argue that the number agreement is not really 'grammatical agreement' but rather aspectual, relating to iteration and distributivity. If one agrees with the notion that the 'aspectual' function of reduplication in Niuean is the primary one, it does seem possible to define the number agreement as one that is actually triggered by the reduplicated verb logically requiring a plural nominal argument, rather than positing the relationship the other way round.

#### 4 Missing links: diachronic/synchronic considerations

Several fundamental questions need to be raised before the description of reduplication in Niuean can be completed. Further research is needed where these lack adequate answers.

- (i) Q: Since reduplication is not fully productive for all lexical roots (or stems), which roots are subject to reduplication and why?
- (ii) Q: Why do some roots undergo the full reduplication process (DUP and reDUP/REdup) and not others, and which roots are subject to REdup and which to reDUP?
- (iii) Q: Why do some reduplication processes only appear in derived forms?
- (iv) Q: Given that in the whole chain of theoretically possible reduplication processes (including within other derivation processes) many are not realised in practice (or have been lost, including the root form), how are we to deal with such synchronic forms (do we admit recourse to diachronic considerations)?

These questions are addressed, though not necessarily resolved, in the following discussion.

- (18) *ahu*<sup>1</sup>, n. smoke ... NO REDUPLICATED FORMS  
*ahu*<sup>2</sup>, n. gall-bladder ... NO REDUPLICATED FORMS  
*ahu*<sup>3</sup>, v.t. to bale, to fetch ...  
*ahuahu*, DUP, v.t. to bale gently ...  
*ahu*<sup>4</sup>, v.t. to slay ... NO REDUPLICATED FORMS

Based on my knowledge of the Niue dictionary corpus I can say that it is predominantly verbs that have reduplicated forms. (But why not all verbs, or at least those which can be subject to 'FREQUENTIVE' and 'intensive' extensions such as *ahu*<sup>4</sup>?) Those from other categories, such as nouns, can be considered special cases. As for verbs, it would require a detailed analysis of the corpus to determine (if possible) exactly which types of verbs have reduplicated forms (and further to ask which of those have which types of reduplicated forms). It appears to me, for example, that verb homonyms will generally differ in their range of reduplicated forms, presumably so as to aid disambiguation.

<sup>4</sup> While the set of examples in (18) is meant to demonstrate the seemingly random nature as to which items get reduplicated, *ahu*<sup>4</sup> is perhaps not a prime example inasmuch as the meaning content may disallow a 'frequentive and intensive' extension, even though a subject or object plural trigger is conceivable. A better example, randomly chosen from the dictionary, would be *huni* v.t. to apply oil, which has no reduplicated forms; if speakers want to express a frequentive and/or intensive meaning extension in this case, they would have to use lexical means, e.g. (lit.) 'he applies oil (as in a massage) every day and very vigorously too'.



- (19) *hele*<sup>1</sup> v.t. to snare ... *fakahелеhele* v.t. to love someone dearly ... \**hehele*,  
 \**helehele*  
*hele*<sup>2</sup> v.t. to cut ... *hehele* v.t. to cut ... *helehele* v.t. to cut into pieces ...  
 \**fakahелеhele*

Such disambiguation may also override phonological rules whereby reduplicated forms of homonyms yield different vowel lengths, as in *gu*<sup>1</sup> v.i. to moan ... *gūgū* v.i. to mumble ... vs *gu*<sup>2</sup> v.i. to respond ... *gugū* v.i. to respond (subject plural).

Of more immediate descriptive importance, however, are those cases where there is no attested unreduplicated form, even though its shape is evident in another reduplicated and/or further derived form, as in (20). There is no attested Niuean word \**moko* meaning 'cold', although there are four homonyms of *moko*, none in any way related to 'cold'. However, we can reasonably establish that in the derived form *mokomia* the root is attested. Hence, it is correct to describe *mokomoko* as DUP (i.e. **not** derived from *momoko*, which is at least historically a REDup). PPN \**moko* confirms such an analysis (in our dictionary work we use Proto Polynesian as 'confirming' evidence, never as 'deciding' evidence).

- (20) *momoko*, v.i. to be cold ... [PPN \**moko* 'cold']  
*mokomoko*, DUP, v.i. to be cool ...  
*mokomia* (*moko-mia*), Suf., v.i. to be affected by cold ...

There are quite a large number of cases where the unreduplicated root form is no longer attested, starting out as it were with the REDup/reDUP form, but where one could argue, on the surface of it, for a phonological (synchronic) explanation. In (21), for example, \**afe* is not attested in any derivation. Given the many examples like (21), as well as the 'supporting' evidence from Proto Polynesian, I suggest that a similar analysis holds here, namely that *aafe* is a historical REDup of \**afe*.

- (21) *aafe*, v.i. to turn, to branch off ... [PPN \**afe* 'deviate, turn aside']  
*afeafe*, DUP, v.i. to branch off repeatedly ...  
*aafeaga* (*aafe-aga*), n. turning point ...

A similar case is illustrated in (22), where we have no root attested in either a derivation or a Proto Polynesian form, although we can be confident that both words are derived via reduplication from a historical base \**aki*.

- (22) *aaki*, v.t. to take out ...  
*akiaki*, (?)DUP, v.t. to take out ... OBJECT PLURAL

In examples where there is no evidence of first mora reduplication, as in (23), we must be cautious about jumping to conclusions. In this case, we have no grounds to claim that the word is derived from a base \**ale*. The protoform suggests that what 'looks' like a reduplication may indeed be a historical root form. Hence we can say that any forms that 'look' reduplicated cannot be said to be so unless there is evidence in the form of an existing unreduplicated form (either free or bounded), plus confirming evidence from Proto Polynesian.

- (23) *aleale*, Qualifier, transparent ... v.i. to be thin ...  
 [PPN \**aleale* 'thin, weak with hunger; hollow or concave']  
*fakaaleale*, Qualifier, thinly ...

With regard to Proto Polynesian forms, which we found very helpful to work with in our dictionary, I can nevertheless demonstrate 'degrees' of usefulness. (In other words, what is

more confounding, the Niuean or the Proto Polynesian data?). The root *aga* in (24) occurs in many more derivations of a fairly diverse semantic nature, so it might have been tempting to include the word *agaaga*, n. spirit, soul ... as a reduplication under *aga*. Yet Proto Polynesian saves us from this course—PPN \*qagaqaga ‘soul, spirit’ points to *agaaga* as a headword of its own.

- (24) *aga*, n. habit, way of acting, behaviour, custom, tradition ...  
 [PPN \*aga ‘habit, custom, way of acting’]  
*fakaagaaga*, Prf., DUP, v.t. to preen oneself ...  
*agaagai*, DUP, Suf., v.t. to surround ...

Native speaker intuition goes against Proto Polynesian in cases (25) and (26), where it is strongly believed that *afiafi* ‘evening’ is derived via reduplication from *afi* ‘fire’ (in the sense that the evening sunset looks like ‘fire’ on the horizon).

- (25) *afi*, n. fire ... [PPN \*afi ‘fire’]  
 (26) *afiafi*, n. evening ... [PPN \*afiafi ‘evening’]

## 5 Conclusion

The semantic and syntactic roles and functions that reduplications have in Niuean—or as I prefer to phrase it, the semantic and syntactic environments that trigger reduplication in Niuean—are broadly similar to the roles and functions generally ascribed to Polynesian languages, namely that of marking ‘FREQUENTIVES’ and ‘plurality’. In detail, however, there is little to which one can compare this present analysis (with the possible exception of Hawaiian), so it remains to be seen whether the other Polynesian languages have a similarly complex array of semantic and syntactic triggers. It is perhaps not surprising that there is a continuum between semantics and syntax, as some reduplications respond to both levels of representation at the same time. Equally unsurprising, perhaps, is that at the syntactic level alone it is mainly a matter of verb–argument agreement, given that at the semantic level alone it is only verbs that respond. Beyond these generalisations the data described yield no rules that allow even for semiproductive derivations. The verbs (and some nouns) that undergo reduplication processes cannot be predicted, nor can the range of reduplication within further derivations. A large-scale corpus analysis would perhaps provide some answers.

## References

- Bybee, Joan L., 1985, *Morphology*. Amsterdam/Philadelphia: John Benjamins.  
 Elbert, Samuel H. and Mary Kawena Pukui, 1979, *Hawaiian grammar*. Honolulu: University of Hawai‘i Press.  
 Krupa, V., 1982, *The Polynesian languages*. London: Routledge & Kegan Paul.  
 Lieber, R., 1992, *Deconstructing morphology*. Chicago: The University of Chicago Press.  
 Marantz, Alec, 1982, Re reduplication. *Linguistic Inquiry* 13:435–482.

- Massam, Diane and Yves Roberge, 1997, Operator-bound clitics and Niuean *ai*. In James R. Black and Virginia Motapanyane, eds *Clitics, pronouns and movement*. Current Issues in Linguistics Theory, 140. Amsterdam: John Benjamins.
- McCarthy, J. and A. Prince, 1995, Prosodic morphology. In J.A. Goldsmith, ed. *The handbook of phonological theory*. Cambridge, Mass.: Blackwell.
- McEwen, J.M., 1970, *Niue dictionary*. Wellington: Department of Maori and Island Affairs.
- Meyerhoff, M. and B. Reynolds, 1996, On reduplication and its effects on the base in Maori. In M. Nespors and N. Smith, eds *Dam phonology: HIL Phonology Papers II*. The Hague: Holland Academic Graphics.
- Mosel, Ulrike and Even Hovdhaugen, 1992, *Samoan reference grammar*. Oslo: Scandinavian University Press.
- Seiter, William, 1980, *Studies in Niuean syntax*. New York: Garland Press.
- Sperlich, Wolfgang B., 1997, *Tohi Vagahau Niue. Niue language dictionary*. Honolulu: University of Hawai'i Press.

