

PACIFIC LINGUISTICS

Series C - No. 88

AUSTRONESIAN LINGUISTICS
AT THE 15TH PACIFIC SCIENCE CONGRESS

Andrew Pawley

Lois Carrington

eds



Department of Linguistics

Research School of Pacific Studies

THE AUSTRALIAN NATIONAL UNIVERSITY

PACIFIC LINGUISTICS is issued through the Linguistic
Circle of Canberra and consists of four series:

SERIES A - Occasional Papers
SERIES B - Monographs
SERIES C - Books
SERIES D - Special Publications

EDITOR: S.A. Wurm

ASSOCIATE EDITORS: D.C. Laycock, C.L. Voorhoeve, D.T. Tryon, T.E. Dutton

EDITORIAL ADVISERS:

B.W. Bender

University of Hawaii

David Bradley

La Trobe University

A. Capell

University of Sydney

Michael G. Clyne

Monash University

S.H. Elbert

University of Hawaii

K.J. Franklin

Summer Institute of Linguistics

W.W. Glover

Summer Institute of Linguistics

G.W. Grace

University of Hawaii

M.A.K. Halliday

University of Sydney

E. Haugen

Harvard University

A. Healey

Summer Institute of Linguistics

L.A. Hercus

Australian National University

Nguyễn Đăng Liêm

University of Hawaii

John Lynch

University of Papua New Guinea

K.A. McElhanon

University of Texas

H.P. McKaughan

University of Hawaii

P. Mühlhäusler

Linacre College, Oxford

G.N. O'Grady

University of Victoria, B.C.

A.K. Pawley

University of Auckland

K.L. Pike University of Michigan;

Summer Institute of Linguistics

E.C. Polomé

University of Texas

Malcolm Ross

University of Papua New Guinea

Gillian Sankoff

University of Pennsylvania

W.A.L. Stokhof National Center for
Language Development, Jakarta;

University of Leiden

B.K. T'sou

Murdoch University;

University of Hong Kong

E.M. Uhlenbeck

University of Leiden

J.W.M. Verhaar

Gonzaga University, Spokane

All correspondence concerning PACIFIC LINGUISTICS, including
orders and subscriptions, should be addressed to:

The Secretary
PACIFIC LINGUISTICS
Department of Linguistics
Research School of Pacific Studies
The Australian National University
Canberra, A.C.T. 2601
Australia.

Copyright © The Authors

First Published 1985

Typeset by Ling Matsay

Printed by A.N.U. Printing Service

Bound by Adriatic Bookbinders Pty Ltd

Maps drawn by Theo Baumann, Cartography, Department of Linguistics, Research
School of Pacific Studies, Australian National University.

The editors are indebted to the Australian National University for assistance in the
production of this series.

This publication was made possible by an initial grant from the Hunter Douglas Fund.

National Library of Australia Card Number and ISBN 0 85883 329 8

TABLE OF CONTENTS

		page
INTRODUCTION		iv
THE CONTRIBUTORS		v
A. RECONSTRUCTION AND SUBGROUPING		
Oceanic subgrouping: retrospect and prospect	George W. GRACE	1-18
Assessing the typological evidence for New Guinea Oceanic	Joel BRADSHAW	19-30
Central-Eastern Oceanic: a subgrouping hypothesis	John LYNCH and D.T. TRYON	31-52
The status of Proto-Micronesian	Byron W. BENDER and Judith W. WANG	53-92
Possessive constructions in Oceanic languages and in Proto-Oceanic	Frantisek LICHTENBERK	93-140
Proto-Oceanic *qi	Robin HOOPER	141-167
Rapid lexical change and aberrant Melanesian languages: Sengseng and its neighbours	Ann CHOWNING	169-198
Languages of north and central Vanuatu: groups, chains, clusters and waves	Ross CLARK	199-236
The status of flora and fauna glosses that have been reconstructed for Proto-Oceanic and for three sub-Oceanic proto-languages	D.S. WALSH	237-256
The position of Atayal in the Austronesian family	Paul Jen-kuei LI	257-280
Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou	Stanley STAROSTA	281-312
B. DICTIONARY MAKING		
The role of computers in dictionary-making at the University of Hawaii	Robert HSU	313-328
The politics of dictionary making on Tanna (Vanuatu)	Lamont LINDSTROM	329-341
Rapid lexical change and the problem of what to include in a new Tahitian dictionary	Jack H. WARD	343-353

INTRODUCTION

The 15th Pacific Science Congress was held in Dunedin, New Zealand, from February 1-11, 1983. Papers submitted from the four Congress symposia dealing with Pacific Island languages are being published by Pacific Linguistics. The present volume contains eleven papers from the symposium "Reconstruction and Classification in the Austronesian Language Family", organised by Robert A. Blust and Andrew Pawley, together with three papers on Austronesian languages from the symposium "Dictionaries for Oceanic Peoples", organised by Lawrence A. Reid.

Papers from the Congress which deal with Papuan languages are to be included in *Papers in New Guinea linguistics No.24 (Pacific Linguistics, A-70)*; papers from the symposium "Pidgin and Creole Languages in the Pacific" are to be published in *Papers in pidgin and creole linguistics No.4 (Pacific Linguistics, A-72)*, whilst those papers to do with Australian linguistics will appear in *Papers in Australian linguistics No.17 (Pacific Linguistics, A-71)*.

The editors wish to thank Stephen Wurm for his advice and guidance, and also, most especially, Ling Matsay, for her thoughtful typesetting.

ANDREW PAWLEY
Department of Anthropology
University of Auckland

LOIS CARRINGTON
Department of Linguistics
Research School of Pacific Studies
Australian National University

THE CONTRIBUTORS

BYRON W. BENDER lived in Micronesia for ten years before joining, in 1963, the faculty of the University of Hawaii, where he is Professor of Linguistics. He continues to work on the languages of Micronesia, and the development of a word and paradigm morphology.

JOEL BRADSHAW's main research interests are in language contact and language change, and the Austronesian languages of New Guinea. He has done fieldwork in the Morobe Province, Papua New Guinea. His address is Apt. 407, 620 Sheridan Street, Honolulu, Hawaii 96814.

ANN CHOWNING is Professor of Anthropology at Victoria University of Wellington, New Zealand. She has carried out anthropological and linguistic research in four Austronesian-speaking societies, three in West New Britain (Lakalai, Sengeng and Kove) and one in Milne Bay (Molima), all in Papua New Guinea. Her interests include comparative linguistics with particular reference to the culture history of Melanesia.

ROSS CLARK is Senior Lecturer in Linguistics in the Department of Anthropology, University of Auckland. In addition to the languages of North and Central Vanuatu, his main research interests are in comparative Polynesian linguistics and language contact.

GEORGE W. GRACE is Professor of Linguistics at the University of Hawaii. His main research interests are historical Austronesian (especially Oceanic) linguistics, linguistic diachrony and the nature of language generally.

ROBIN HOOPER is completing a doctorate in linguistics in the Department of Anthropology, University of Auckland. Her research is concerned with comparative Polynesian grammar and the Tokelau language. She is an editor and advisor to the Tokelau Dictionary project.

ROBERT W. HSU is Assistant Professor of Linguistics at the University of Hawaii. His main research areas are in computer applications in language data processing, the languages of Micronesia and phonology.

PAUL JEN-KUEI LI is Research Fellow and Chairman of the Linguistics Section, Institute of History and Philology, Academia Sinica, Taipei. He is interested in the Austronesian and Sino-Tibetan language families and has done extensive fieldwork on Formosan languages.

FRANTISEK LICHTENBERK is Senior Lecturer in Linguistics at the University of Auckland. His research interests are in Oceanic descriptive linguistics, Oceanic historical syntax and language typology.

LAMONT LINDSTROM is Assistant Professor of Anthropology at the University of Tulsa, Oklahoma. His main field of research has been the island of Tanna (Vanuatu). In addition to work on kava and a dictionary of the Kwamera language of the south-eastern coast of the island, he has recently completed a book, *Knowledge and power in a South Pacific society*, which applies post-structuralist theory to island discourse.

JOHN LYNCH is Professor of Language at the University of Papua New Guinea. His main research interests are in descriptive and comparative studies of the languages of Melanesia, particularly those of Southern Vanuatu.

STANLEY STAROSTA is Professor of Linguistics at the University of Hawaii. He has done fieldwork on Austronesian languages in Taiwan and on Munda languages in South India. His main theoretical interests are in syntax.

DARRELL TRYON is a Senior Fellow in Linguistics in the Research School of Pacific Studies, Australian National University. His main research interests are in the Austronesian languages of Oceania. He has done fieldwork in the Loyalty Islands, the Solomons, Vanuatu and Tahiti.

DAVID WALSH is Senior Lecturer in Anthropological Linguistics in the Anthropology Department, University of Sydney. His research interests are in Oceanic languages (lexicography, historically oriented comparison, ethno-taxonomy) and in Pacific pidgins and creoles.

JUDE WANG is completing a doctorate in linguistics at the University of Hawaii. Her main research interests are in Oceanic linguistics (especially Micronesian), artificial intelligence and computers.

JACK H. WARD is Associate Professor in the Department of Indo-Pacific Languages at the University of Hawaii. He has done fieldwork in Indonesia, Tahiti and the Tuamotus, and is currently compiling a dictionary of Tahitian.

OCEANIC SUBGROUPING: RETROSPECT AND PROSPECT

George W. Grace

PROLOGUE

It was in 1953, thirty years ago this year, that I first began to work in Austronesian linguistics and in fact to work on the problem of Oceanic subgrouping. There were only a handful of people working in the field at that time. When I look around today, I see big changes. I see first of all a great increase in the number of people actively working in the field. How large the number has become was brought home to me when I momentarily considered the idea of reading here¹ a list of the younger people in the field today whose work deserves recognition. Almost as soon as I started drawing up the list, I gave up the idea. What I quickly discovered was that there are simply too many people who are making truly significant contributions to contemplate listing them.² In fact, it is very gratifying to reflect on the current state of the field. The linguistic descriptions and all of the historical and other interpretations being made seem to be based on much more adequate information than in those early days. They also seem more sophisticated and, in some important respects, more insightful. So maybe it sounds as if what I should say here is simply that the present generation is already in the process of doing the things I would have wanted to do and doing all of them much better than I would have been able to do them, and that I, therefore, gratefully would now pass them the torch. And so saying, presumably I would sit down.

Unfortunately, perhaps, it is my understanding that that is not what is expected of me — that I am expected to fill more time than that. Consequently, I have prepared a paper to deliver to you.

What I will talk about in this paper is an idea, which I have had for a long time, that there are special conditions affecting Oceanic, and particularly Melanesian, languages. Whatever these conditions are, I felt that they were not adequately foreseen in the assumptions, explicit and tacit, of comparative-historical linguistics — the assumptions of what I have called our 'general theory of change and differentiation'. As a result, we were, I thought, largely left to our own devices in learning to understand them. But, it seemed that if and when we ever did discover what was involved, our discoveries would have implications for some of the assumptions of linguistic theory in general. I want here to develop this idea further. I will attempt to analyse some of our problems and make some suggestions as to how our assumptions about linguistic change and differentiation might be modified.

* * * * *

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 1-18. *Pacific Linguistics*, C-88, 1985.

© George W. Grace

1

It always seemed to me that we who were working in the Oceanic field had much more difficulty in solving our subgrouping problems than linguists working in other areas. I can remember being asked by other linguists for a quick overview of the Austronesian family or of Oceanic, just pointing out the main branches, and so forth. I, of course, was forced to reply that those questions were still very much a matter of dispute, that we regarded answering questions such as those as a quite long-range objective. The response always seemed to be something like, "Well, you've had people working out there for some time, there must be *something* you can tell us by now". And I'd wind up telling them about our controversies. Or more exactly our one main controversy, because there was one controversy of overriding importance. That one concerned the origin of the non-Polynesian Austronesian languages of Melanesia — what are generally known as the 'Melanesian languages'.

In the words of Wilhelm Milke in 1962 (Milke 1962), there were two competing schools of thought on the problem. One school assumed that these languages, along with the Polynesian languages and most of the languages of Micronesia, derived from a single proto-language. That is, it assumed that they constituted a subgroup of the Austronesian family, a subgroup now usually known as 'Oceanic'. The set of assumptions which characterise this school may be referred to as the 'Oceanic hypothesis'. The second school assumed that small groups of Indonesians had established colonies in many parts of Melanesia and that pidginised languages had developed in these sites. The Melanesian languages, according to this school, are derived from these pidginised Indonesian languages. This set of assumptions can be referred to as the 'pidginisation hypothesis'. However, as if the situation were not already bad enough, at the very moment when Milke was writing his comment on the two schools, a third one was coming into being. This school, based on Isidore Dyen's (cf. Dyen 1962, 1965) lexicostatistical classification of the Austronesian languages, assumes that the Melanesian languages actually belong to a number of different subgroups of Austronesian, and, in fact, that Melanesia is probably the original homeland of the Austronesian family. I will refer to this set of assumptions as the 'Melanesian homeland hypothesis'.

An observer today would probably note that most of those linguists currently concerned with the historical linguistics of Oceania appear to have accepted the Oceanic hypothesis, and he or she might conclude from that observation that the differences which formerly separated the three schools have been satisfactorily resolved. Now, I have always believed the Oceanic hypothesis to be essentially correct, and I have no intention of abandoning it today, but I have always been troubled by the realisation that the other schools were both designed to explain certain facts, or at least observations which appeared to reflect facts, and that some of these observations are not explained by the Oceanic hypothesis. What has happened is that we have simply set them aside, at least for the time being.³

For some years I have felt increasingly uncomfortable about the facts which we have not yet been able to explain, and increasingly I have felt that the explanations will continue to elude us as long as we do not alter our way of looking at these facts. A little later I will be referring to some articles dealing with classification in the biological sciences, but I would like to jump ahead here and interject a quotation from one of them which I find particularly apt in the present connection. The article from which I am quoting is an influential one by Niles Eldredge and Stephen Jay Gould. Early in the article they make the following statement (Eldredge and Gould 1972:83): "The expectations of theory color perception to such a degree that new notions

seldom arise from facts collected under the influence of old pictures of the world. New pictures must cast their influence before facts can be seen in different perspective".

To bring this back to Oceanic linguistic classification, I think that traditional comparative historical linguistics does have its own picture of the world. I think that the school in Oceanic linguistics which espouses the Oceanic hypothesis generally falls within this orthodox tradition. And I think that to a great extent those facts which we have been unable to explain are facts which indeed were not collected under the influence of our picture of the world. They were collected by adherents of the other schools, and at least some of these qualify as non-believers in our orthodox tradition. And the reason why we cannot explain these facts is that they do not fit into our picture — there is no proper place in the picture for them — they are alien facts.

That is why I think we need a new picture. We need a picture that will accommodate all of the known facts. I have come to doubt that our reconstructions of the history of changes and differentiations can ever be very accurate or very complete unless we modify our picture of the processes involved. It is my purpose here to suggest some modifications. In so doing I will be referring to ideas from the biological sciences and from sociolinguistics.

One particular point at issue in Oceanic linguistics has been the possibility of language mixture. On this point, Arthur Capell, who is the dean of living advocates of the pidginisation hypothesis, has further clarified his position in an article published in 1976 (Capell 1976). I would like to quote one statement from that article. He says (1976:529), "Even a so-called 'mixed language' has a single parent-language. But such a language also has an invader, and when the balance of relationship to the original mother tongue — the true ancestor — is outweighed by the invader's contribution in lexicon and structure, it is right to speak of a mixed language." At another point (1976:528-529) he uses the following metaphor: "But a language may also be 'taken over' by another like a wall whose mortar is eaten out by ivy and held together, so far as it is, in a new way."

Capell's statements suggest that we might find some reconciliation of views if we were prepared to acknowledge that a language might, although classifiable on the basis of its ancestry as a member of a certain genetic grouping, still have undergone changes of sufficiently fundamental nature as no longer to have the characteristics which one would expect to find in languages of that group. This last point is somewhat difficult to discuss, because we are not accustomed to talking about it except informally.

The point is that genetic linguistic groupings, particularly where the relationship is close, tend to have the character of what we might call 'typological classes' as well. For example, it seems perfectly reasonable to speak of a Polynesian linguistic type, and to say that all of the present day Polynesian languages in fact conform to this type. The explanation for their all being of the same type is quite obvious of course — they are all continuations of the same comparatively recent proto-language, which is to say that they are closely related genetically. There is nothing in the least surprising about this likeness; it could hardly be otherwise. And of course it is because of these typological similarities that knowing one language in such a group is of great assistance when one seeks to analyse another. Such similarities have an important place in our picture of the world. If we are told that something is a Polynesian language or a Romance language, we immediately have a rather

complex and detailed idea of what the language is like. And if the language should turn out to be different from what we expect — different to a really serious extent — then we may feel virtually betrayed. We may be tempted to assert that the language does not deserve to be called 'Polynesian' or whatever it is.

What I understand Capell to be saying is that if we have a language which is genetically a member of one group but has many of the typologic characteristics of another group, then the language has not been fully classified unless both of these facts have been given recognition. To look at it from a slightly different angle, we might say that one concern which underlies the notion of mixed languages is that the *unshared* innovations which have occurred in a language may have a significance which should not be overlooked. I will say more about unshared innovations later.

Now, I find it very interesting to observe that classification in botany and zoology has been faced with something of the same problem. In fact, taxonomy in biology presents a number of the same problems which we encounter in attempting to determine linguistic relationships. The close analogies between the genealogical models used in linguistic and in biological classification have frequently been pointed out. Each focuses upon the developments affecting one particular kind of unit — in linguistics, the individual language, and in biology the species. Each of these units in its own field plays a crucial role as the entity which is supposed to maintain its identity over time while still undergoing changes. The species or the language can, of course, from time to time split into several descendant species or languages, but distinct species or languages can never *merge*. Once established, they maintain their integrity until they become extinct.

In a recent article in *Science* entitled "Biological classification: toward a synthesis of opposing methodologies", Ernst Mayr (1981) discusses three competing methodologies in biological classification. With a bit of forcing I could match each of these with one of the three schools in Oceanic historical linguistics, but I do not think the results justify the effort. What I want to discuss here is the differences between just two of them. These two he calls respectively, 'cladistics' [a designation taken from the Greek word for 'branch'] and 'evolutionary classification'. Evolutionary classification is classification in the Darwinian tradition — or more precisely in a neo-Darwinian 'modern synthesis' which was proclaimed in the 1950s (cf. Gould 1982). Cladistics (which is often called by its advocates 'phylogenetic systematics') is usually dated from the publication in 1950 of Willi Hennig's *Grundzüge einer Theorie der phylogenetischen Systematik* (Hennig 1950). The aims and methods of cladistics are very analogous to those of genetic classification in linguistics. Their objective is to reconstruct the order of branchings which have led to the known taxa; in sum, to construct the correct family tree. Their principal evidence is shared innovations (called 'synapomorphies').⁴

The principal criticism made of this school by the evolutionary classificationists is that they are concerned only with changes which are evidence for branching — i.e., with shared innovations. They are not concerned with whatever other changes have occurred. This disagreement seems to parallel a major disagreement between the advocates of the pidginisation hypothesis, who accord an important role to language mixture, and those of the Oceanic hypothesis, who do not. The Oceanic hypothesis is based on the objective of reconstructing the order of branching, and shared innovations constitute its critical evidence. The pidginisation hypothesis (among other things) attaches great importance to

what has happened to a language since it split off from its closest genealogical relatives; it is therefore concerned also with unshared innovations.

To get back to these problems as they confront the biological systematists, one of the consequences of adopting a cladistic taxonomy would be, for example, that crocodilians (crocodiles, alligators, and the like) would be put into a branch with the birds which would exclude all of the other reptiles. The crocodilians and the birds are descended from the stem group archosaurians, which is not the ancestor of turtles, snakes, and lizards. In fact, in a cladistic taxonomy there could not be any class of reptiles which excluded either birds or mammals. The class Reptilia constitutes what is called a *paraphyletic* taxon — i.e., a category which includes some but not all of the descendants of a particular ancestor. The fishes, as a group exclusive of their descendants the amphibians, reptiles, birds, and mammals, also constitute a paraphyletic taxon as do the pongidae — a grouping composed of the orang utans, gorillas, and chimpanzees, but excluding man.

A mixed language as characterised by Capell would seem to be one case where unshared innovations and paraphyletic groupings might be accorded a place in linguistics. The unshared innovations would be those which led to the language's qualifying as mixed in the first place. And if we should decide that because of its radical changes it no longer qualified as an Austronesian language, for example, or as an Oceanic language, then Austronesian or Oceanic so defined as to exclude it would be a paraphyletic taxon.

However, while we are considering these notions we should inquire whether or not it is only in clear cases of language contact that they might have useful applications. We might ask whether or not it is possible that some of our languages are birds, so to speak, and some reptiles. And we might ask, in case it is possible, what the implications might be. For example, we might ask whether, if we were faced with such a case, we would come to the cladistically proper phyletic tree or whether we would be likely unbeknownst to ourselves to set up a paraphyletic grouping instead — a paraphyletic grouping which would then be sailing under false colours.

How, we may ask, would our birds appear to us? One important difference between our situation and that of biology is that we have no concept comparable to that of evolutionary grades. In biology there are perfectly reasonable grounds for saying that birds and mammals represent *progress* beyond the reptilian grade. When a paraphyletic group such as the class Reptilia is set up, it is natural to expect those descendants which have been excluded (e.g., birds and mammals from the Reptilia) to be not only more radically changed, but also to be evolutionarily more advanced. Would our bird-like language appear in any way more advanced than the reptilian grade languages from which it had sprung? My guess is that it would not, that we might even be tempted to regard it with a trace of disapproval.

But the idea of evolutionary progress is not necessary in order to justify paraphyletic groupings. The crucial fact that justifies such groupings in biology is the fact that some evolutionary lines (e.g., the birds) have changed in fundamental ways whereas during the same time their sister lines (e.g., the crocodilians) have undergone only relatively insignificant changes. That the changes made by the birds are regarded as constituting evolutionary progress is incidental. The analogy could be a productive one for linguistics if linguistic change also proceeds unevenly. But does it?

The evidence seems at first glance to be conflicting. On the one hand we have those languages which are candidates for mixed language status, which I have been describing as characterised by extensive unshared innovations. In fact, the rate of change in English during the two centuries or so following the Norman conquest seems to have been particularly rapid. But even so, languages which have been subject to intensive contact might simply be a special, essentially abnormal, case. However, it also seems that some languages are more conservative overall than their sisters. Fijian, among Oceanic languages, has been characterised as particularly conservative, and the rate of change in Icelandic during the last 700 to 800 years seems to have been especially slow (cf. Bergsland and Vogt 1962).⁵ I think the conclusion should be that, although the profession has not given its blessing to the idea that rates of linguistic change can vary significantly, there is reason to suspect that that is the case.

On the other hand, our professional paradigm seems to assume that most linguistic change is caused by conditions internal to the language itself. That would tend to suggest that the rate and degree of change are largely unaffected by the environment in which the language is spoken. And, although that does not necessarily require that change must proceed at a constant rate, I believe that it at least *suggests* that the rate will be *relatively* constant. In short, I think that we have tended at least implicitly to hold a gradualistic picture of the processes of change.

I want to say something in a moment about models of change in biology. But first, I would like briefly to consider the implications which it might have for Oceanic linguistics if it is true that some languages are more conservative than others. Most obviously, we might expect that there would be some linguistic analog of paraphyletic taxa. That is, we might find languages which seemed to group together on the basis of shared retentions, that seemed to group together as opposed to some of their sisters because these sisters had changed so significantly through unshared innovations that they seemed to have extracted themselves from the group. Are there, in fact, any candidates for paraphyletic group status? And, if so, how have they been treated in our linguistic classifications?

We might think first of all of the notion 'Melanesian language'. Our cladistic classifications do not recognise a Melanesian group which would exclude the Polynesian and most Micronesian languages, and yet there are some traits which are widely distributed in Melanesia but absent in Micronesia and Polynesia. Perhaps the best example is the possessive systems, where there is a system which seems typical of Melanesia generally from New Guinea to Fiji, while the characteristic Polynesian system is quite different, and the characteristic Nuclear Micronesian system also seems to have a quite different character now. However, we do not, most of us, on the basis of such considerations recognise Melanesian as a linguistic grouping. If we found it useful to recognise paraphyletic groupings for some purpose, we might then see fit to set up a Melanesian paraphyletic grouping.

Are there any other examples? One immediately thinks of the Malayo-polynesian Linkage of Isidore Dyen's 1965 lexicostatistical classification. Lexicostatistics in principle makes no distinction between shared innovations and retentions, and, in fact, I believe that most of Dyen's higher level groupings would turn out to be based in the main on relatively large numbers of shared retentions. The Malayopolynesian Linkage, then, would be another example of a paraphyletic grouping, but it, of course, is also not generally accepted as a valid grouping from the cladistic point of view.

Now we come to more questionable, but also more interesting instances. For example, there seem to be several cases in Melanesia where some speakers of one language in a chain of closely related languages have migrated elsewhere and, as a result of subsequent changes, their language has become more differentiated from its stay-at-home sister languages than the latter are from their geographical neighbours. It has been suggested (cf. Geraghty 1983) that Polynesian has migrated from eastern Fiji with such results, and likewise Rotuman from Fiji (cf. Pawley 1979) and Central Pacific (Fijian, Polynesian, and Rotuman) from central Vanuatu (Pawley 1977). It is not clear, however, whether the migrating language has undergone more rapid than normal change or whether those which remained at home have undergone change at a slower than normal rate or even whether the latter have changed at the same rate but shared some of their changes. In any case, if we recognise a grouping consisting of the stay-at-home sister plus its neighbours — e.g., a grouping consisting of all Fijian dialects or one consisting of all of the languages of northern and central Vanuatu — that grouping would be a paraphyletic one.

I wonder if Eastern Oceanic in its original conception might not be such a paraphyletic grouping. The original Eastern Oceanic group was defined as consisting of Fijian, Polynesian, Rotuman, the languages of central and northern Vanuatu, and those of most of the south-eastern Solomon Islands. Geographically, this grouping would, with some exceptions, include all of the Austronesian languages of an area beginning in the extreme east of Polynesia and extending westward to embrace the eastern section of Melanesia. The exceptions which I mentioned are all in eastern Melanesia. They are the Austronesian languages of the Santa Cruz Islands, southern Vanuatu, the Loyalty Islands, and New Caledonia. To suggest that Eastern Oceanic might be a paraphyletic grouping is to suggest that some of the languages which are in fact descended from its last common ancestor have changed so radically that they no longer have the same general character as the proto-language and its more conservative descendants. If there are descendants of the Eastern Oceanic ancestral language which have become so radically changed, it is likely that they are to be found among the languages which are spoken in eastern Melanesia but which have been excluded from Eastern Oceanic.

I believe that I will have no difficulty in getting your consent to the general proposition that none of the languages of eastern Melanesia which have been excluded from Eastern Oceanic would be thought of as conservative languages. You would probably also consent to the proposition that those languages which were included in Eastern Oceanic generally *could* be characterised as conservative. However, I feel that I would probably be well advised not to push the analogy to reptiles vs. birds and mammals too far. I will not go so far as to suggest that the languages of the Santa Cruz Islands, southern Vanuatu, the Loyalty Islands, and New Caledonia have, analogously to birds and mammals, evolved to a higher grade than the languages recognised as Eastern Oceanic.

I have said, then, that I believe that there is good reason to accept the general principle that some languages do change more rapidly than others. However, as I mentioned above, our general conception of the processes of change would tend to suggest that change should be gradual and relatively constant. In biology also the accepted model of change was until recently a gradualistic one. However, in recent years that model has been challenged. In 1972 Niles Eldredge and Stephen Jay Gould published the paper from which I quoted earlier. It was entitled, "Punctuated equilibrium: an alternative

to phyletic gradualism" (Eldredge and Gould 1972). As the title indicates, they proposed a new model for change — particularly the change which leads to new species. In so doing they argued that the reason that the gradualistic model had survived so long in biology was not that it had the support of the evidence. It was that biologists assumed it to be true, and therefore managed to see the evidence as being compatible with it.

The same point might be applicable to any tacit assumption of gradualism in linguistic change. Any claim about what anyone's tacit assumptions are is very vulnerable to challenge, of course. Let me say, at least, that I myself have certainly tended to make assumptions which were surely informed by such a tacit assumption of gradualism. For example, I have regularly assumed that the degree of similarity between different languages is a fairly reliable basis for judging the relative length of the periods of their independent development.

At any rate I would like to say a few words about the proposed new model of biological change, called the 'punctuated equilibrium model'; I would then like to consider the possible implications for linguistics if such a model of change should prove to be applicable to language.

The punctuated equilibrium theory has recently received what has been interpreted as very strong empirical support in a much discussed report by Peter G. Williamson which appeared in *Nature* in 1981. The report, which deals with the evolution of a number of mollusc species in the vicinity of Lake Turkana in Kenya during the late Cenozoic, has been extensively referred to. The importance of this work rests on the remarkably complete fossil record for a period of several million years found in floodplain and delta deposits there. In his report, Williamson (1981:437) makes the unequivocal statement that, "Evolutionary patterns in all lineages at the site conform to the 'punctuated equilibrium' model: no 'gradualistic' morphological trends occur." I want to sketch out some of the features of the punctuated equilibrium model as presented by Eldredge and Gould and by Williamson's work. According to this model:

1. Species within their ancestral environments normally quickly reach a homeostatic equilibrium in which there is strong resistance to any change.⁶

2. The periods in which significant change occurs are geologically very short — i.e., very short in relation to the periods of stasis. It typically occurs in a population of the species which exists in its own particular microenvironment — probably a stressful one for them.

3. The population which undergoes the rapid changes is usually isolated from the rest of its conspecifics in an environment outside the main range of the species.⁷

4. The period of change is marked by an unusually high level of phenotypic variance — that is, the individual members of the group undergoing change show much more variation among themselves than is normally to be found in a population in equilibrium.

Let me take up these points one by one and discuss their possible application to linguistic change. The first point was that species within their ancestral environments normally quickly reach a homeostatic equilibrium in which there is strong resistance to any change. This may seem particularly foreign to us linguists at first glance because we are accustomed to looking for the cause of linguistic changes in disequilibrium within the affected language. Since there is assumed always to be some change going on, we are accustomed to thinking of all languages as being constantly in disequilibrium at least to a slight degree. However, there is a reverse to that coin, namely that it is precisely the assumption that the language tends to seek equilibrium

that makes it possible to conceive of *disequilibrium* as a cause of change. Moreover, the very fact that we assume that the points of disequilibrium can be specified, indicates that such points are pictured as minor irregularities in a near equilibrium. Presumably, therefore, it does in fact follow naturally from our assumptions that a language is ordinarily quite resistant to random changes.

The second point was that the periods in which significant change occurs are geologically very short — i.e., very short in relation to the periods of stasis — and that it typically occurs in a population of the species which exists in a particular microenvironment which is probably a stressful one for them. This seems to be more difficult for linguistics to accommodate. The theory of punctuated equilibria suggests that a species has a fairly clearly marked beginning and end in time as well as a bounded distribution in space. Furthermore, a new species may exist simultaneously with the species from which it evolved. In fact, if one species invades the range of the other, they may come into direct confrontation.

There seem to be several problems in attempting to apply this aspect of the model to language. First, we are not accustomed to considering how languages are adapted to their external environments at all, and therefore it is hard to know just what might constitute significant differences in microenvironments and what environmental conditions might be stressful. However, there does seem to be one kind of phenomenon in linguistics which fits the description — the development of pidgin (or creole?) languages. To take a specific example, if we regard Tok Pisin as being a genetic continuation of English, then we do seem to have a case where a new language developed very rapidly out of another while the other continued to exist. Of course, Tok Pisin is only one of many examples of the pidginisation process.

But there is more to be said. It is, I think, becoming increasingly apparent to sociolinguistics that creole languages are not so much a special kind of phenomenon as the limiting cases of a quite common one. The common phenomenon is what William Labov (1965:110-112) calls "change from below" (the level of consciousness). The relevant environmental factor, I believe is multilingualism on the part of some of the people who speak the affected language (not necessarily its native speakers). Such a situation is often referred to as 'language contact', but that term is misleading. It suggests that one language in some way acts directly upon the other.

I would like to suggest that what is really significant in so-called 'language contact' situations is the fact that some of the utterances in the language in question are produced by people who, at least part of the time, *think* in another language. By 'thinking in a language' I mean roughly what I understand the lay public to mean by it. I mean formulating utterances directly in that language rather than formulating them first in another and then translating them. People who are thinking in a certain language may be said to experience reality in terms of that language, or at the very least to *interpret* their experiences in its terms. Anyway, I propose that if a language is sometimes spoken by people who have done some of their thinking in another language, that fact is a significant factor in its environment.⁸ A language must be under pressure to acquire the means to express the thoughts of those who speak it, and it is thus that one language influences another. This leads naturally, I suggest, to change from below, and, if that persists long enough, it can eventually result in the condition which Capell has described.

Thus, other languages can be significant factors in the environment of a particular language. However, I would not want to suggest that they are the *most important* environmental factors, but only that their effects are relatively easy to recognise. I assume that a language encounters its external environment, in the guise of the uses to which it is required to lend itself. For example, it will presumably experience some pressure in terms of what its speakers want to use it to talk about and how. That, in turn, will to a considerable extent be a reflection of the general physical and cultural environment in which the language is spoken.⁹

So much for point 2. The third point was that the population which undergoes the rapid changes is usually isolated from the rest of its conspecifics in an environment outside the main range of the species. I have, in fact, heard it suggested that languages spoken by small, isolated populations change more rapidly, but this point seems also to be one on which the paradigm of linguistics has taken no position. Some evidence which might seem to indicate that it is true may be found in the cases which were mentioned above of Polynesian, isolated from Fiji; and Central Pacific, isolated from central Vanuatu.

Point 4 was that the period of change is marked by an unusually high level of phenotypic variance — that is, the individual members of the group undergoing change show much more variation among themselves than is normally to be found in a population in equilibrium. What would be the appropriate linguistic analog of phenotypic variance? I would like tentatively to suggest that individual utterances are the analog of the individual phenotypes in a biological population. With this interpretation the hypothesis would be that languages in a state of accelerated change would be characterised by a relatively high degree of permissiveness as regards standards of correctness. This involves both the claim that speakers of some languages take a more authoritarian position concerning correctness in their language than do others and the claim that a less authoritarian stance is indicative of accelerated change in progress. The first seems to be an empirical matter, but one not usually reported on by field investigators. However, there are cases elsewhere in the world where standards of correctness have been reported as particularly lax. Examples are the northern Athabaskans of North America (as reported by Scollon and Scollon 1979:206, for example)¹⁰ and the Paliyan of India as reported by Gardner (1966). I am not sure what kind of evidence might bear on the second claim — that rapid change occurs in such circumstances — but it seems a priori to be reasonable.

CONCLUSIONS

I believe that the above discussion indicates that some of the phenomena which have been observed in biology do have their counterparts in linguistics. This suggests, therefore, that we might have something to gain by modifying what Eldredge and Gould call our "picture of the world" so as to incorporate some of these new points. It seems to me that to do so would help to fit the facts brought to light by each of the schools of thought in Oceanic linguistics into the same picture. I would hope that it would also provide a basis for the discovery of new facts which we might otherwise have overlooked.

I do not want to suggest that all aspects of the new picture are entirely clear at this point. In fact, there will surely always be room for further specification of details. However, some aspects do seem clear from what has been said here. To begin with, I believe that we must conclude that languages do not change at a constant rate.

Once this assumption is put into the picture, other aspects of the picture adjust accordingly. In fact, acknowledging that languages change at uneven rates seems to have implications of two sorts for the rest of the picture. First there are implications for interpreting our data — implications about what kind of data would be produced as a result of uneven linguistic change. In particular, we must now reckon with the possibility that, of any group of sister languages — i.e., languages descended from the same proto-language — some may have undergone significantly more and more fundamental changes than others. And we must be concerned about the implications which follow from that possibility.

Thus we might expect to find within any set of sister languages some which have been more conservative, and hence are more like the proto-language. These more conservative languages will presumably be much more useful in reconstructing the proto-language. They also will probably be much more like one another than like their more innovative sisters. They are likely, therefore, to tempt us to group them together — a paraphyletic grouping, a sort of linguistic class *Reptilia*. The implication for our own practices of this assumption and of the experiences of biological systematics seems to be that we should pay still greater attention to the distinction between what I have called typological criteria and the criteria by which genetic branchings are reconstructed.

Another way in which our picture of the world is likely to be affected by our recognising change as proceeding unevenly lies in what that conclusion implies about the causes of change. Why would it be that languages do not change at a constant rate, or at least at a relatively constant rate? It seems to me that what this suggests is that the causal factors themselves are not very constant. That, in turn, suggests (although it could certainly not be said to prove it) that causal factors external to the system — causal factors which do not arise within the language itself — are relatively more important than we have imagined them to be. To attempt to construct a theory of linguistic change in which the explanation of the changes is located within the language itself has been very appealing to linguists because this permits them to remain on familiar ground throughout. They need deal only with phenomena within their professional domain of competence.

However, I think it is becoming increasingly apparent that languages change also on the basis of the uses to which they are put. It has been suggested that revolutionary political changes such as have occurred in this century in Russia and in China have had significant effects on the respective languages. A substantial literature is developing dealing with the effects of literacy on languages. A probably related phenomenon is the movement in many countries of the so-called 'third world' to 'modernise' their languages. Finally, there is the effect of so-called 'contact' with another language — which I have tried to convince you is also at bottom a matter of the uses to which the language is put.

Now, what would be the consequences of accepting this revised picture of the world as I have described it? Well, first of all, I think as I said above that some of the difficulties and frustrations which we have encountered, particularly in dealing with some of the Melanesian languages, may contain lessons for the linguistic world at large.

But what is the source of these difficulties and frustrations? Why should we have had such difficulty reaching agreement on even the most basic questions about the linguistic relationships in Oceania? Well, for one thing, it is true that the linguists who came to work in the area came from quite different backgrounds, and that fact no doubt contributed to the diversity of their viewpoints.

But it is also true, I believe, that most of our problems are not the artifacts of our backgrounds but come from the linguistic conditions in the area. But what, our detractors might ask, is exceptional about these linguistic conditions? Well, I do not have a complete answer to that question, but I would suggest that it has something to do with the fact that a large proportion of the Melanesians live in linguistic boundary areas — that is, they live close to a boundary between the territory of one language and that of another. In fact, it has been pointed out by various observers that the areas in which many of the Melanesian languages are spoken are so limited in size as to be, in effect, boundary area throughout.¹¹ In most parts of the world linguists have tended to regard the linguistic varieties spoken in boundary areas as abnormal, and it is only recently that such varieties have begun to receive much attention. Moreover, even then the attention has come mainly, not from historical linguists, but from sociolinguists.

What I would suggest, then, is that the assumptions on which comparative historical linguistics rests were developed under conditions where the linguistic varieties spoken in boundary areas could be ignored. In Melanesia they cannot, and for that reason I suggest that Melanesia provides a particularly valuable test of the universality of these assumptions. Consequently, I would suggest that the new picture which I have tried to sketch out here, a picture which arises most directly from our struggles with the phenomena of our own languages, will eventually have to be acknowledged as valid for diachronic linguistics generally.

EPILOGUE

There is one point on which I would like to expand a bit further in this written version than I did in the original address. That point is the need to distinguish more meticulously the criteria by which genetic branchings are reconstructed from those which I have called typological. As I understand it, it was this distinction that was brought into sharper focus in biology by the advocates of the cladistic approach. In linguistics, our classifications purport to represent genetic branchings and our reconstructions to represent nodes in a genetic tree. Interested representatives of other disciplines such as archaeology are generally encouraged to understand them in this way. However, it seems to me that we do sometimes permit ourselves to use the paraphernalia of this cladistic model — i.e., the family tree and the concept of reconstruction — in ways which do not strictly conform to cladistic objectives. There are two kinds of such instance which I have in mind.

First, I believe that in the drawing of family trees languages are sometimes excluded from a particular subgroup largely on the negative argument that there is little or no evidence that suggests that they belong to it. However, careful practice would require that their exclusion from the group should be regarded as indeterminate until *positive* evidence of their proper position can be found. (What I am saying is of course not intended to imply that there is anything at all improper in making *working assumptions* which exclude them as long as the provisional nature of the assumptions is clearly recognised). I must express appreciation to Ross Clark in his contribution to this Congress for having made the basis of the positive argument clear to me. He pointed out that what counts as evidence that a language is *not* a member of a particular subgroup is that it retains from a still earlier stage features which have not been retained by the members of the subgroup in question.

Thus, for so long as it cannot be established in the case of a particular Austronesian language either that it shares the innovations which characterise a particular subgroup or that it retains features which were not retained by the languages of that subgroup, its position in relation to that subgroup must be regarded as indeterminate.

Second, I believe that the concept 'reconstruction' is sometimes used loosely for procedures which, although more or less analogous to those of comparative historical reconstruction, are being applied without much concern for what connection the reconstructed system might have with historical reality. Such 'reconstructions', I believe, are sometimes used by linguists who have no particular interest in history as a device for providing a unified theory for (some component of) some assortment of languages. That is, such reconstruction provides a sort of theory for a set of what I am calling typological similarities without providing any clear statement of what the theory is intended to explain or how. However, it leaves the inexplicit suggestion that some sort of historical explanation is intended.

NOTES

1. This is a slightly revised version of the keynote address delivered to the symposium on Reconstruction and Classification in the Austronesian Language Family at the 15th Pacific Science Congress, Dunedin, New Zealand, 1-11 February 1983.
2. Just to give an idea of the kinds of contributions I was thinking of I will give a few indications, but I would like to emphasise that the list is intended only to be indicative. To begin with, there has been the great impetus to syntactic reconstruction provided by Andrew Pawley and his work. Also there have been a continuing series of contributions to historical reconstruction in Oceanic by Robert Blust, Malcolm Ross, Paul Geraghty, Sheldon Harrison, Richard Levy, Frantisek Lichtenberk, and others. There has been progress on the understanding of language contact in the area from the work of Tom Dutton and Joel Bradshaw among others. Furthermore, descriptive treatment of the languages has progressed gratifyingly in the last few years. Work on what I have sometimes called the 'aberrant' languages has made great progress in recent years. We have for the first time a group of dedicated and experienced linguists working on the difficult languages of New Caledonia and the Loyalty Islands (I would like to mention Jean-Claude Rivierre, Françoise Ozanne-Rivierre, Jacqueline de la Fontinelle, and Claire Moyse-Faurie), and the work of John Lynch has radically improved our knowledge of the also difficult languages of southern Vanuatu. Some work has been done on the Austronesian languages of the Santa Cruz Islands. Darrell Tryon has completed surveys of Vanuatu and of the Solomon Islands. Much has been added to our knowledge of the Micronesian languages which, a few years ago, were notoriously poorly known. These contributions have been made mainly by scholars at the University of Hawaii and have been due largely to the initiative of Byron Bender. Also to be mentioned are the Proto-Polynesian reconstructions of Bruce Biggs and his collaborators. And it should be emphasised that this list makes no pretence to being complete.

3. The observation upon which the Melanesian homeland and the pidginisation hypotheses are based can be described roughly as follows: That on which the Melanesian homeland hypothesis is based is that in Dyen's lexico-statistical study many Melanesian languages showed only very low cognate percentages with all other Austronesian languages.

The perceptions which led to the pidginisation hypothesis seem to have been roughly the following (I base this account largely upon Arthur Capell 1962 and Sidney Ray 1926):

(1) That there is a characteristic Melanesian grammatical structure which is very similar to the Micronesian structure, "departs a little more widely" (Capell 1962:387) from the Polynesian, and "diverges very widely" from the characteristic grammatical structures of the languages of Indonesia.

(2) That the Melanesian grammatical structure appears simplified in relation to the Indonesian.

(3) That the vocabulary items which are shared by Melanesian languages other than immediate neighbours are relatively few in number and that almost all of them have cognates in Indonesia.

(4) That some Melanesian languages possess more of these vocabulary items than others and that the items have been less reshaped by phonological changes in some languages than in others.

(5) That those Melanesian languages which seem best to have preserved the original vocabulary and grammar are generally those spoken on the smaller islands.

(6) That some Melanesian languages show various Papuan-like features. A particularly conspicuous example is OV word order in the New Guinea area generally (cf. Capell 1969). There are a number of changes in particular languages such as those of Santa Cruz and the Reef Islands (Lincoln 1978, Wurm 1978), Maisin, and Magori (cf. Dutton 1976).

4. Other terms employed in cladistics include: apomorphy (=innovation), autapomorphy (=unshared innovation), plesiomorphy (=retention), symplesiomorphy (=shared retention).
5. About Icelandic, Bergsland and Vogt say (1962:128), "[In the case of Icelandic], the extensive borrowings from German and Danish found in the Norwegian Riksmål have not disturbed the linguistic conservatism of a patriarchal society with an almost millennial literary tradition, which remains unbroken to the point that children, as soon as they learn to read at all, read without difficulty the prose written by their forefathers some 700 or 800 years ago."
6. Incidentally, this equilibrium is not based on gene flow (in fact, Williamson found no difference in the patterns for sexual and asexual species) but apparently upon the complex interaction of genes with genes. Thus, the adaptation of each gene to its genetic environment becomes a stabilising factor.
7. At least this is the claim that has been made. I do not feel that Williamson's data support it. The claim is, in effect, the so-called 'allopatric' theory of speciation. As I understand it, when that theory was originally formulated, isolation of the affected population from normal gene flow was considered to be a particularly significant factor. Later, Ernst Mayr (Mayr 1954) suggested that interrupted gene flow was not so important. He suggested that allopatric speciation was due in large part to what came to be called the 'founder effect' — the fact that the

small population migrating to a new environment would bring with it only an incomplete sample of the original gene pool. However, the instances of speciation found by Williamson involve asexual as well as sexual species — which seems to eliminate gene flow as a critical factor, and involve quite large populations which have remained in place — which seems to rule out any very important role for the founder effect. The key causal factor seems to have been climatic changes which led to environmental stressing.

8. I should qualify this remark by pointing out that in some places in Melanesia neighbouring languages seem to be mutually so closely adapted that thinking in one is almost identical to thinking in the other. In such a case, the presence of the other language might have very little effect; cf. note 11 below.
9. There is one potential source of confusion which should be mentioned. We have become accustomed to referring to the speakers of a language as its 'speech community'. However, recent work in sociolinguistics compels us to acknowledge that the X-language speech-community is probably not, in fact, a community. The speakers of a language may not all belong to the same community, and the members of an actual community may speak more than one language. In fact, patterns of behaviour involving the use of different languages may be well integrated into the life of the community.

As a consequence, the analogy between species and languages may be misleading here. The species must adapt to the general environment in which its members live. Of course, there must also be linguistic adaptation to the environment — in this case cultural as much as physical — of the community. However, the entity which must be adaptive to this environment is not the language per se except in the limiting case where the community uses only one language; it is the linguistic repertoire of the community. The total linguistic resources of the community must, then, presumably be suitably adapted to the general physical and cultural environment. However, if we want to follow the established pattern of focusing upon the individual language as the unit to be observed over time, we must consider other languages in the repertoire of the community or communities which speak it to be part of the effective environment.
10. Regarding the Athabaskans, the Scollons say (Scollon and Scollon 1979:206): "To the individualistic and integrative bush consciousness, language is also highly integrated to the personal experience of the speaker. There are no normative standards or higher values placed on the speech of others. One is one's own expert, but at the same time because of the nonintervening aspect, no attempt is made to impose one's own values on others. Thus language is seen as highly individualized and variable."
11. Languages — what we call languages — are to a large extent emblematic entities. Although this is not a familiar situation in the modern Western World, there are many places in areas such as Melanesia where there are a number of different languages spoken within a narrow geographical scope, but where all of them show essentially the same adaptation to the same physico-cultural environment. That is, they are functionally interchangeable. The choice among them is of no significance except for what it says about the speaker's interpretation of him/herself and of the situation in which he/she is speaking. The situation which I am describing is that which André Haudricourt (1961) has called "egalitarian bilingualism". What we are dealing with I think is a bilingualism where the only

functional difference between the languages is the emblematic one. Maurice Leenhardt (1930, 1946) emphasised the great similarities in the conceptual systems of New Caledonia languages, pointing out that for a speaker of one New Caledonian language learning another one or translating into another one was an entirely different matter from learning French or translating into it.

BIBLIOGRAPHY

BERGSLAND, Knut and Hans VOGT

1962 On the validity of glottochronology. *Current Anthropology* 3:115-153.

CAPELL, Arthur

1962 Oceanic linguistics today. *Current Anthropology* 3:371-428.

1969 *A survey of New Guinea languages*. Sydney: Sydney University Press.

1976 Austronesian and Papuan "mixed" languages: general remarks. In S.A. Wurm, ed. *New Guinea languages and language study*, vol.2: *Austronesian languages*. PL, C-39:527-579.

CLARK, Ross

1983 Subgrouping of central and northern Vanuatu. Paper presented to the 15th Pacific Science Congress, Dunedin.

DUTTON, T.E.

1976 Magori and similar languages of south-east Papua. In Wurm, ed. 1976:581-636.

DYEN, Isidore

1962 The lexicostatistical classification of the Malayopolynesian languages. *Language* 38:38-46.

1965 *A lexicostatistical classification of the Austronesian languages*. Indiana University Publications in Anthropology and Linguistics. *International Journal of American Linguistics*, Memoir 19.

ELDREDGE, Niles and Stephen Jay GOULD

1972 Punctuated equilibria: an alternative to phyletic gradualism. In Thomas J.M. Schopf, ed. *Models in paleobiology*, 82-115. San Francisco: Freeman, Cooper.

GARDNER, Peter M.

1966 Symmetric respect and memorate knowledge: the structure and ecology of individualistic culture. *Southwestern Journal of Anthropology* 22:389-415.

GERAGHTY, Paul

- 1983 The history of the Fijian languages. *Oceanic Linguistics Special Publication no.19*. Honolulu.

GOULD, Stephen Jay

- 1982 Darwinism and the expansion of evolutionary theory. *Science* 216 (23 April):380-387.

HAUDRICOURT, André G.

- 1961 Richesse en phonèmes et richesse en locuteurs. *L'Homme* 1:5-10.

HENNIG, Willi

- 1950 *Grundzüge einer Theorie der phylogenetischen Systematik*. Berlin: Deutscher Zentralverlag.

LABOV, William

- 1965 On the mechanism of linguistic change. *Georgetown University Monograph Series on Languages and Linguistics* 18:91-114.

LEENHARDT, Maurice

- 1930 *Notes d'ethnologie néo-calédonienne*. Travaux et Mémoires de l'Institut d'Ethnologie, 8. Paris.
- 1946 *Langues et dialectes de l'Austro-Mélanésie*. Travaux et Mémoires de l'Institut d'Ethnologie, 46. Paris.

LINCOLN, Peter C.

- 1978 Reef-Santa Cruz as Austronesian. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 929-967. PL, C-61.

MAYR, Ernst

- 1954 Change of genetic environment and evolution. In Julian Huxley, A.C. Hardy, and E.B. Ford, eds *Evolution as a process*, 157-180. London: George Allen and Unwin.
- 1981 Biological classification: toward a synthesis of opposing methodologies. *Science* 214 (30 Oct.):510-516.

MILKE, Wilhelm

- 1962 Comment on Arthur Capell, Oceanic linguistics today. *Current Anthropology* 3:415-416.

PAWLEY, Andrew

- 1977 On redefining 'Eastern Oceanic' (draft mimeo).
- 1979 *New evidence on the position of Rotuman*. Working papers in Anthropology, Archaeology, Linguistics, and Maori Studies. Department of Anthropology, University of Auckland.

RAY, Sidney H.

- 1926 *A comparative study of the Melanesian island languages*. Cambridge: Cambridge University Press.

SCOLLON, Ronald and Suzanne B.K. SCOLLON

1979 *Linguistic convergence: an ethnography of speaking at Fort
Chipewyan, Alberta.* New York: Academic Press.

WILLIAMSON, Peter G.

1981 Paleontological documentation of speciation in Cenozoic molluscs
from Turkana Basin. *Nature* 293:437-443.

WURM, Stephen

1978 Reef-Santa Cruz: Austronesian, but ...! In Wurm and Carrington,
eds 1978:969-1010.

ASSESSING THE TYPOLOGICAL EVIDENCE FOR NEW GUINEA OCEANIC

Joel Bradshaw

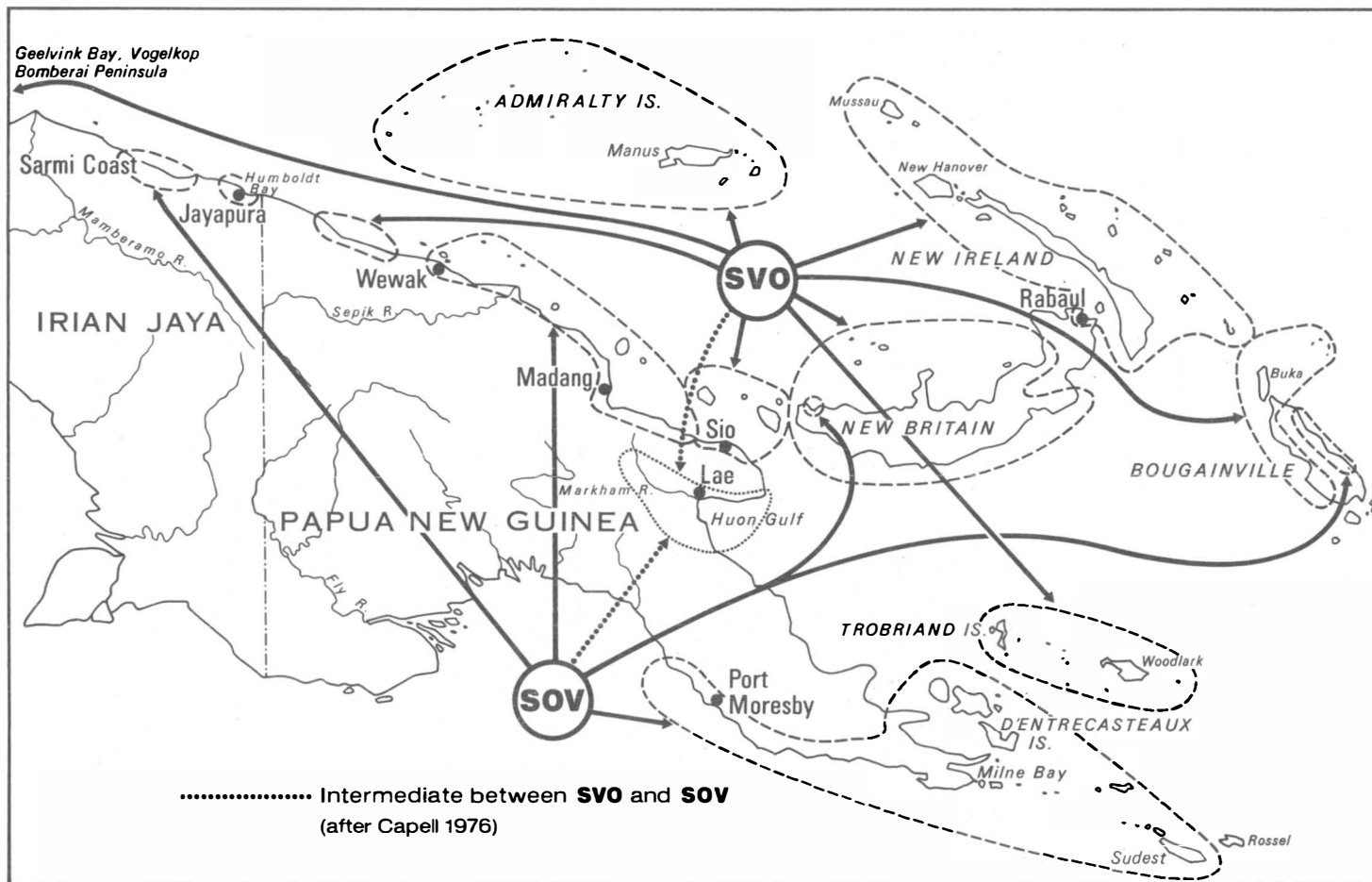
The Oceanic languages on or adjacent to the New Guinea mainland exhibit in common some striking typological innovations. They are the only group of languages in the Austronesian family that possess a significant number of features usually associated with OV word order. The unique typology of New Guinea Austronesian languages is almost certainly innovative and apparently results from contact with non-Austronesian languages on the New Guinea mainland, which are nearly universally OV.

The languages exhibiting the most markedly innovative traits are confined to the eastern half of the New Guinea mainland and adjacent offshore islands. The common Austronesian VO pattern is found elsewhere, in the area sometimes referred to as 'Island Melanesia' (New Britain, New Ireland, Manus, and elsewhere). The Morobe languages in the north-eastern part of the New Guinea mainland fall somewhere between these two extremes. The distribution of word order typologies in the Papua New Guinea area is summarised in Table 1 and on the accompanying map.

Table 1			
Distribution of word order types in Papua New Guinea			
Common Austronesian, Island Melanesia & elsewhere:			
VO	Prepositions	N + Gen	N + Adj
PNG Austronesian, Morobe Province:			
SVOV	Prepositions Postpositions Ambipositions	Gen + N	N + Adj
PNG Austronesian, elsewhere on the mainland:			
SOVV	Postpositions	Gen + N	N + Adj
Common Papuan (PNG non-Austronesian):			
SOV	Postpositions	Gen + N	N + Adj

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 19-30. Pacific Linguistics, C-88, 1985.

© Joel Bradshaw



Capell's typological classification of Austronesian languages in Papua New Guinea

For some time, researchers have hoped to establish that this typologically distinctive group is a genetic subgroup. The hypothesis is that some or all of the changes that distinguish these languages occurred only once in a language ancestral to all of them. This is now known as the New Guinea Oceanic hypothesis, and the hoped-for subgroup generally goes by the name 'New Guinea Oceanic'. (It is generally agreed that the languages in question belong in the Oceanic subgroup of the Austronesian language family.) But, so far, no solid basis for uniting the whole group has been found. Since I need to refer to the member languages of this group, I will continue to use this name. However, in my usage 'New Guinea Oceanic' labels a typological unit, not a genetic one.

Some of the weaknesses of the evidence for a genetic grouping follow:

(1) The lexicostatistical evidence indicates greater *disunity* among New Guinea Oceanic languages than among practically any other regional group of Austronesian languages (see Dyen 1965).

(2) The phonological evidence for a New Guinea Oceanic subgroup is limited to the merger of Proto-Oceanic *d and *R (Milke 1965:343; Pawley 1978:143). Since this merger has occurred in many other Oceanic languages, it is hardly a firm base for a subgroup.

(3) Lexical evidence is more abundant, but it is relatively weak evidence for methodological reasons. When we isolate a group of languages with uniquely shared vocabulary, we can rarely be sure that the items uniquely shared are actually innovations within the group. They may be retentions from a proto-language ancestral to languages outside as well as inside the group. Low-level reconstructions have a habit of rising to higher and higher levels as more evidence is brought to bear on them. Free-standing words are also notoriously easy to borrow, and old borrowings are not often easy to identify.

(4) Grammatical evidence is the only remaining hope of those who still see possibilities for a New Guinea Oceanic subgroup (see Chowning 1973:226; Milke 1965:346; Pawley 1978:134-141). Good grammatical evidence may yet turn up. However, there are two major weaknesses with most of the kinds of evidence available.

Some of the grammatical evidence either fails or is weakened because it simply lacks the proper distribution. Chowning (1973) and Pawley (1978) have already faulted most of Milke's (1965) grammatical evidence on distributional grounds. For instance, preposed genitives and reinforced possessives are too abundantly distributed. They are found in non-Oceanic Austronesian languages of New Guinea and eastern Indonesia as well as in strictly New Guinea Oceanic languages. Classificatory prefixes, on the other hand, are too sparsely distributed. They are absent in the languages of Central Papua and of Morobe Province. Pawley's (1978:138) favourite piece of grammatical evidence — the use of reflexes of Proto-Oceanic *i-ai as a locative postposition — can also be faulted on the same grounds. To my knowledge, no reflexes of *i-ai show up in languages west of Morobe Province. In its place, some languages west of Morobe show a generalised locative postposition that appears to derive from a postposed, possessed noun originally meaning 'inside'. (In Gedaged, this postposition takes the shape -lon; in Manam, -lo.)

The focus of this paper, however, is not the distribution of grammatical evidence, but rather the methodological problems posed by the typological nature of much of the potential grammatical evidence. The methodological problem arises as one attempts to reconstruct the means by which innovative morphology is acquired. Most of the innovative morphology that distinguishes

New Guinea Oceanic from other Oceanic languages results from the use of common Oceanic morphemes in innovative positions or innovative functions. To the extent that these positional and functional innovations are easily acquired and attributable to a change of environment, their usefulness as evidence of genetic affiliation diminishes. Features easily acquired are suspect because they can easily be borrowed and are apt to have changed more recently. Structural changes in response to a change of external environment are suspect because they are not unexpected and thus do not require heredity as an explanation. They tell more about environment than heredity. (Heredity, of course, may tell us a lot about much more ancient environmental adaptation.)

When Oceanic languages first arrived in the New Guinea area, they left an environment in which VO word order was the norm. They encountered a new environment in which OV word order was the norm. The grosser aspects of OV word order acquired by all New Guinea Oceanic languages do not suffice to prove that all share a common ancestor apart from other Oceanic languages. Nor does the innovative morphology that derives in unsurprising fashion from those grosser word-order innovations. The crucial elements in subgrouping are the more surprising and unexplainable innovations.

In some ways, the uniqueness of the New Guinea Oceanic languages within the Austronesian family is analogous to the uniqueness of marine mammals within the class of mammals. It is a mistake, for instance, to argue for a close genetic relationship between whales, on the one hand, and sea lions, on the other, just because both have adapted in similar ways to an environment radically different from the usual mammalian surroundings. Cetaceans and pinnipeds are united by certain shared differences from the common run of mammals. Both groups lack coarse, outer body hair, have flippers instead of external limbs, and are streamlined for swimming. But these shared traits are important for typological, not genetic, classification. The characteristics most crucial for determining the heritage of creatures like whales and sea lions are those that are hardest to predict. Whales, for instance, propel themselves with what used to be their tails and steer with what used to be their forelegs. Sea lions propel themselves with their erstwhile forelegs and steer with their former hindlegs. This indicates that protocetaceans and protopinnipeds were somewhat different creatures when each began adapting to a marine environment.

To take another example from mammalian biology, hairiness and size do not provide crucial evidence in determining genetic affiliation. One need only think of the enormous variation within the dog species. In historical and comparative linguistics, word order does not provide crucial evidence in subgrouping. The order of words in a language (that is, the order of relatively contentful, free morphemes) is relatively easy to tamper with and the limits of variation are quite narrow. One syntactic category may either precede or follow another syntactic category. (The only other option — infixing one word within another — is quite abnormal.) Chances are good, then, that unrelated languages with similar syntactic categories will share many similarities of word order. Japanese, Korean, and Finnish, for instance, share many word-order traits with Indic and Dravidian languages. Thus the well-attested innovations in word order shared by many New Guinea Oceanic languages are not very powerful evidence for subgrouping, especially when these word-order innovations look like adaptations to the word order usually encountered in the non-Austronesian languages of New Guinea.

Unlike word-order innovations, shared innovations in grammatical morphology are usually considered the next best thing to shared phonological innovations as a basis for subgrouping. The position of grammatical morphemes (that is, of relatively contentless, bound morphemes) is relatively immutable. Moreover, grammatical morphemes are prone to exhibit certain peculiarities of form and function that make the presence of similar forms performing similar functions in separate languages a striking fact. The presence of a causative prefix with a shape derivable from *pa- or *paka- in so many Austronesian languages is just such a striking fact. Table 2 lists examples of this prefix in widely distributed languages. Chances are good that languages exhibiting such similarities shared a period of common development and derive from a common ancestor.

Table 2	
Austronesian causatives	
RUKAI, Formosa (Li 1973:70)	
'a-'acay	kuani taraalu' sa babuy
<i>cause-die</i>	<i>that hunter ART boar</i>
	<i>that hunter killed a boar</i>
ILOKANO, the Philippines (Lawrence Reid, p.c.)	
im-pa-kan	na diay baboy
<i>GF-cause-eat</i>	<i>3s that pig</i>
	<i>he fed the pig</i>
MALAGASY, Madagascar (Simeon Rajaona, p.c.)	
n-amp-anasa	ny lamba aho
<i>PAST-cause-wash</i>	<i>the clothes 1s</i>
	<i>I had the clothes washed</i>
ROVIANA, the Solomon Islands (Todd 1978:1039)	
va-mate-a	sa si keke boko pa inevana
<i>cause-die-3s</i>	<i>3s PART a pig for feast</i>
	<i>he killed a pig for the feast</i>
BAUAN, Fiji (Apenisa Seduadua, p.c.)	
eratou vaka-mate-a	na vuaka
<i>3p cause-die-3s</i>	<i>ART pig</i>
	<i>they killed the pig</i>
HAWAIIAN, the Hawaiian Islands (Bill Wilson, p.c.)	
ho'o-make	lākou i ka pua'a
<i>cause-die</i>	<i>3p OBJ ART pig</i>
	<i>they killed the pig</i>

But what happens when separate but related languages sharing a similar stock of morphemes and a similar grammar begin to rearrange their word order in similar ways in response to similar pressures? Relatively contentful, free morphemes will begin to appear in innovative positions. These initial changes will be externally motivated and perhaps rather abruptly and disruptively accomplished. Over time, some of the rearranged morphemes will evolve into relatively contentless, bound morphemes performing functions required by the new word order. Grammatical morphemes already in useful positions may take on additional duties required by the new word order. This second phase of change will be internally motivated and probably much more gradual. During this stage, the innovations are assimilated and nativised. Eventually, chances are good that the shared innovations in word order will produce shared innovations in grammatical morphology. The languages in the same environment will thus show common morphological innovations without being especially closely related to each other. The shared morphological innovations will be the by-products of changes which do *not* constitute good evidence of genetic affiliation. The development of a new word order will be relatively easily accomplished and will be based on external linguistic models. The development of various corollaries of the new word order will be based partly on external models, and partly on internal requirements for efficient parsing. The tendency for words performing grammatical functions to lose content and freedom over time is perhaps one of the gradual sorts of erosion that all languages are subject to.

The history of the classificatory prefixes of many New Guinea Oceanic languages illustrates the derivation of 'good' morphological evidence from 'bad' syntactic evidence.

In 1943, Capell listed sets of verbal prefixes from the Austronesian languages of South-East Papua. These prefixes are now called 'classificatory' because they classify into a limited number of categories the manner in which an action is performed (action by hitting, cutting, holding, etc.). The prefixes often occur attached to noun and adjective roots as well as verb roots. Some of the prefixes are transparently related to full verbs in the same language (to hit, to cut, to hold, etc.). In such cases, the classificatory prefix constructions resemble straightforward verbal compounds. Other prefixes have so little semantic content that their original meanings are irrecoverable without comparative evidence. The role of the almost contentless prefixes in many languages is hard to separate from that of the common Austronesian causative prefix. Table 3 compares the role of the inherited causative prefix in Hawaiian with the role of some of the innovative causative constructions in New Guinea Oceanic languages.

In 1965, Milke pointed out correspondences between certain classificatory prefixes in Gedaged and certain ones in South-East Papuan languages. He proposed that the classificatory prefixes were a morphological innovation providing evidence for a New Guinea Oceanic subgroup. However, since there are no such prefixes in the Morobe Province languages in the middle of the north coast, nor in the Central Province languages on the south coast, this piece of evidence lacks the proper distribution to tie together all the New Guinea Oceanic languages.

Table 3	
Some Hawaiian and New Guinea Oceanic causative constructions	
HAWAIIAN	ho'o-hana (<i>work</i>) to employ, cause to work
WEDAUI	rau-karäi (<i>work</i>) to set (s.o.) to work
HAWAIIAN	ho'o-helele'i (<i>falling</i>) to scatter, sow
WEDAUI	ravi-awawari (<i>falling</i>) to sow broadcast
HAWAIIAN	ho'o-hua (<i>fruit</i>) to bear fruit
NUMBAMI	-ambi ano (<i>fruit</i>) to bear fruit
HAWAIIAN	ho'o-huli (<i>turn over</i>) to turn, change, convert
IWAL	-amb nalili (<i>turned around</i>) to turn (s.t.) around
HAWAIIAN	ho'o-loli (<i>turn, change</i>) to change, amend
NUMBAMI	-ambi lele (<i>turned</i>) to translate
HAWAIIAN	ho'o-luli (<i>shake</i>) to rock (s.o.); to sway
WEDAUI	ravi-dagudagu (<i>restless</i>) to shake, disturb
HAWAIIAN	ho'o-make (<i>die</i>) to kill, let die
MANAM	rau-mate (<i>die</i>) to kill
HAWAIIAN	ho'o-piha (<i>full</i>) to fill
TUBETUBE	ro-karapowani (<i>full</i>) to fill
HAWAIIAN	ho'o-pi'i (<i>ascend</i>) to cause to rise
GEDAGED	bi-sa (<i>ascend</i>) to lift up, raise
HAWAIIAN	ho'o-puka (<i>perforation</i>) to make a hole or opening
NUMBAMI	-so bozoka (<i>punctured</i>) to make a hole or opening

There are no classificatory prefixes in the VO languages of Morobe Province. But there are main verbs which play a role similar to that of the prefixes. These main verbs cooccur with verbal, nominal, and adjectival roots whose semantics resemble those of the verbs, nouns, and adjectives occurring with classificatory prefixes in the OV languages. Some of the morphemes involved are cognate as well. (The forms meaning *to hit* and *to cut* have widely distributed cognates.) Table 4 shows some VO constructions in Morobe Province languages which correspond to OV constructions in languages in other parts of Papua New Guinea.

Table 4					
Reflexes of the serial causative in New Guinea Oceanic languages					
* S	V _t	O	V _i	(Switch S)	<i>they hit pig die</i>
* S	V _t	O	V _t	(Same S)	<i>they hit pig kill/cause-die</i>
The VO languages:					
S	V	O	V	GITUA (Switch S)	ti-rap ngaya mate 3p-hit pig 3s-die
S	V	O	R	NUMBAMI (fr Same S)	ti-lapa bolo uni 3p-hit pig dead
The OV languages:					
S	O	V	V	KAIRIRU (Switch S)	bur rro-un-i a-myat pig 3p-hit-3s 3s-die
S	O		V-V	GEDAGED (fr Same S)	boz du-punu-fun-i pig 3-p-shoot-kill-3s
S	O	clas-V		MANAM (fr Switch S)	boro di-rau-mate-i pig 3p-hit-die-3s
S	O	clas-V		IDUNA (fr Same S)	bawe hi-lu-ve-'alika-na pig 3p-*hit-cause-die-3s

There are thus innovative causative constructions in both VO and OV languages which resemble each other in the semantics of the components involved, in the order in which the components occur, and, in many cases, in the shapes of the individual morphemes as well. Moreover, the two groups of innovative causatives (OV and VO) are in complementary distribution in Papua New Guinea, and both differ from the causative pattern commonly found in Oceanic languages elsewhere. These circumstances seem to justify the reconstruction of a single ancestral pattern that will account for both the VO and OV constructions. A central element of this pattern, the SVOV serial causative, is reconstructed in Table 4.

Two types of SVOV serial causative are widely reconstructible. In the Same Subject type, both verbs are transitive and both have the same subject and object. In the Switch Subject type, the second verb is intransitive, and the object of the first verb is the subject of the second. In both constructions, the first verb describes a manner of action and the second a resulting state. Table 5 shows that reflexes of both types can be reconstructed in the same language.

Reflexes of the serial causative are found in a much larger proportion of New Guinea Oceanic languages than the classificatory prefixes are. (They are still lacking in the OV languages of Central Province.) The classificatory prefixes are just one of the various reflexes of the serial causative. But, although the parent construction has a better distribution, it is a much weaker kind of evidence. The classificatory prefixes used to look like an arbitrary

Table 5	
Reflexes of two types of serial causative in Manam	
SWITCH SUBJECT: * S s-V-o O=S s-V yields S O s-V-V-o	(di) boro di-rau-mate-i 3p pig 3p-hit-die-3s <i>they killed the pig</i>
The second verb is intransitive:	boro i-mate pig 3s-die <i>the pig died/is dead</i>
SAME SUBJECT: * S s-V-o O s-V-o yields S O s-V-V-o	(ŋai) 'ai i-zan-sere'-i 3s wood 3s-punch-split <i>he split the stick (end-on)</i>
The second verb is transitive:	(ŋai) 'ai i-sere'-i 3s wood 3s-split-3s <i>he split the wood</i>

morphological innovation until they were tied into the larger pattern. They did not correspond directly to structures in neighbouring non-Austronesian languages. The serial causative, on the other hand, is a word-order innovation which could have come about in a relatively short period of time. Moreover, it appears to be a direct adaptation to the OV word order of the non-Austronesian languages. Table 6 shows how SVOV serial causatives can be parsed ambiguously depending on whether VO or OV word order is considered the norm, and on whether the construction is thought to consist of one clause or two. The meaning remains the same no matter how the structure is parsed syntactically. I suspect that the use of this syntactically (but not semantically) ambiguous construction may have been one of the first steps on the path toward OV word order.

Table 6		
Parsing options for SVOV serial causatives		
	Compatible with SVO	Compatible with SOV
Two clauses:		
Switch Subject	S V O ## V _i	S V ## S V _i
Same Subject	S V O ## V _t	S V ## O V _t
One clause	S V _{main} O V _{sub}	S V _{sub} O V _{main}

It appears that speakers of New Guinea Oceanic languages at some point considered both verbs to occur in a single clause since in most languages either the manner verb or the result verb has lost its verbal status and become a member of a more restricted grammatical paradigm. Table 7 shows this development. To me, this suggests that the serial causative was, from its inception, essentially one clause with a verb in both medial and final position.

Table 7	
Different resolutions for two verbs in one clause	
The VO Solution, adopted by Numbami and other VO languages:	
S V _{manner} O V _{result}	becomes S V _{main} O Result
The OV Solution, adopted by Manam and other OV languages:	
S V _{manner} O V _{result}	becomes S O Manner-V _{main}

The initial development of the serial causative need not have taken place in only one ancestral language. It could have arisen independently whenever individual Austronesian languages came under pressure to move toward the OV word order of their Papuan neighbour languages. It is the type of change that can be accomplished relatively easily and abruptly. The gradual conversion of that innovative two-verb syntactic construction into a one-verb clause containing innovative morphology or grammatical classes no doubt took a much longer time to accomplish. But it too needs no unique genetic explanation. The tendency for new syntax to produce new morphology over time must exert constant, even if not absolute, pressure on every language.

Unfortunately for the New Guinea Oceanic hypothesis, much of the grammatical evidence available to support it is weak for the same reason the serial causative and its end-products are weak evidence for subgrouping. In order to subgroup New Guinea Oceanic languages, we need to identify those morphological innovations which show that a particular group of languages chose one set of options instead of other possible alternatives in response to similar environmental pressures. It is not enough to show that all acquired certain general traits of OV word order and their various morphological end-products. In order to unify the whole group, we need to show that, although different alternatives were available, all chose at least one set of options in common. In response to generalised external pressure to create a locative postposition, for instance, we need to show that all chose one means rather than another to create one (say, the recruitment of the anaphoric locative pronoun *i-ai rather than the semantic extension of a common postposed locational noun like 'inside'). If all the possible alternatives that we can identify are equally distributed throughout the group we want to unify, then chances would seem greater that individual languages chose from the same 'universal' set of alternatives independently. It would be very helpful to have a control group of languages which chose a different set of options when faced with environmental pressures similar to those facing New Guinea Oceanic languages. Until we find such a control group,

or at least until we otherwise filter out innovations which are broadly predictable from changes in the external environment or from universal principles of language change, 'New Guinea Oceanic' will have to remain a typological rather than a genetic label.

NOTE

I wish to thank Frank Lichtenberk for reading this paper on my behalf at the 15th Pacific Science Congress. I also wish to thank George Grace for profitable discussions and Andy Pawley for detailed comments on the written version. This paper elaborates on some of the implications of the evidence cited in my dissertation (Bradshaw 1982). All the data I cite are from that work.

BIBLIOGRAPHY

BRADSHAW, Joel

- 1982 Word order change in Papua New Guinea Austronesian languages. Ph.D. dissertation, University of Hawaii.

CAPELL, Arthur

- 1943 *The linguistic position of south-eastern Papua*. Sydney: Australasian Medical Publishing Co.
- 1976 Features of Austronesian languages in the New Guinea area in general in contrast with other Austronesian languages of Melanesia. In: S.A. Wurm, ed. *New Guinea area languages and language study*, vol.2: *Austronesian languages*, 235-282. PL, C-39.

CHOWNING, Ann

- 1973 Milke's 'New Guinea cluster': the evidence from northwest New Britain. *Oceanic Linguistics* 12:189-243.

DYEN, Isidore

- 1965 *A lexicostatistical classification of the Austronesian languages*. *International Journal of American Linguistics*, Memoir 19.

LI, Paul Jen-kuei

- 1973 *Rukai structure*. Institute of History and Philology, Special Publication No. 64. Taipei: Academia Sinica.

MILKE, Wilhelm

- 1965 Comparative notes on the Austronesian languages of New Guinea. *Lingua* 14:330-348.

PAWLEY, Andrew K.

1978 The New Guinea Oceanic hypothesis. *Kivung* 11:99-151.

TODD, Evelyn M.

1978 Roviana syntax. In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings, fascicle 2: Eastern Austronesian*, 1035-1042. *PL*, C-61.

CENTRAL-EASTERN OCEANIC: A SUBGROUPING HYPOTHESIS

John Lynch and D.T. Tryon

This paper represents the first report on a long-term project which has as its major aim a subgrouping of the Oceanic languages, based largely on morphological evidence. Previous classifications of the Oceanic subgroup of Austronesian give the impression that Proto-Oceanic split into a large number of first-order branches.¹

1. PREVIOUS ATTEMPTS AT LARGE-SCALE SUBGROUPING

Since the recognition of an Oceanic subgroup by Dempwolff, there have been few major attempts at an overall subgrouping of the Oceanic languages, largely because of the huge number of languages involved, and the unavailability of representative data.² Grace published a tentative subgrouping of Oceanic, which he subsequently revised in 1971 and again most recently in 1981.³ His original classification divided Oceanic into nineteen subgroups, but it seems clear that Grace did not necessarily intend that these should be interpreted as first-order subgroups. The original nineteen subgroups were distributed geographically as follows: ten in Papua New Guinea, four in the Solomons, four in New Caledonia and the Loyalties, and one comprising the languages of Vanuatu, Fiji, Rotuma, Polynesia, as well as those languages known as Micronesian. The two later versions of the paper show certain changes in details, mainly as more data became available from regions not well known to Grace in the 1950s. Thus we find, for example, that in the 1971 revision Grace adds two branches in Irian Jaya, while in the 1981 paper he recognises more groups in the Papua New Guinea region. Although various subgroups proposed by other scholars are mentioned in this latest revision, it is not always clear whether Grace is happy to accept them. Because of this, and also because of the format of the 1981 paper, it is difficult to see just exactly how many subgroups Grace would now recognise.

The other major classification was that of Dyen, who published his lexicostatistical classification of the whole Austronesian family, which included material from both Oceanic and western languages, in 1965. This classification in fact found no lexicostatistical basis at all for an Oceanic subgroup. Instead, Dyen proposed a forty-branch family-tree for Austronesian, with thirty-three of these branches being wholly or partly located in the region which most other linguists believe is occupied by the Oceanic subgroup. One of these

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 31-52. *Pacific Linguistics*, C-88, 1985
© John Lynch & D.T. Tryon

thirty-three branches, the large Malayopolynesian Linkage, had most of its members in the western (Indonesian) region, but also included the languages of Polynesia, Rotuma, and Fiji, as well as Mota and the languages of Efate in Vanuatu, Kerebutu, Lau, and To'ambaita in the Solomons, and Motu in Papua New Guinea. The other thirty-two branches had no such western connections; and although two of them — Yapese and Nauruan — may not in fact belong to the Oceanic subgroup, the remaining thirty branches are clearly recognised to be Oceanic by the vast majority of Austronesianists, and were distributed geographically as follows: two in Irian Jaya, seventeen in Papua New Guinea, three in the Solomons, four in Vanuatu, three in New Caledonia, and one in Micronesia.

Despite the considerable differences in approach and results, these two attempts at large-scale classification clearly have one feature in common: no large subgroups are recognised, and any 'family-trees' of Oceanic that could be drawn from these classifications must of necessity be unstratified and show many first-order branches. A considerable amount of more recent research has attempted to draw together some of the subgroups recognised by Dyen or Grace and to combine them into somewhat larger groups: we might mention here in particular research into the Austronesian languages of the New Guinea/Western Solomons area by Blust, Bradshaw, Johnston, Lincoln, Ross, and Tryon, among others. However, the bases on which these proposals have been made differ, and most of them do not at this stage seem to have gained general acceptance.

The project we are engaged on will eventually involve comparison of grammatical morphemes and some grammatical structures across as wide a range of Oceanic languages for which data are available. This present paper discusses the first results from this project. We will show in section 3 that there is a body of apparent morphological innovations indicating a subgroup which expands the Eastern Oceanic (EO) group as defined by Biggs (1965) and Pawley (1972). Apart from including the Micronesian languages (Grace 1955), the expanded subgroup also includes the Southern Vanuatu subgroup and the languages of Utupua and Vanikoro in the Santa Cruz area. We call this putative subgroup Central-Eastern Oceanic. Before examining the evidence for the Central-Eastern Oceanic subgroup however, we first discuss the various hypotheses relating to the Eastern Oceanic group.

2. EASTERN OCEANIC

The term 'Eastern Oceanic' was apparently first used by Biggs (1965). In his classic article on the historical phonology of Rotuman, Biggs adopted as a working hypothesis the existence of an EO subgroup:

... there is some evidence that Fijian, Polynesian, Rotuman and certain languages of the Solomons-New Hebrides chain ... are members of a single subgroup of Austronesian. This paper assumes such to be the case (Biggs 1965:383).

Nothing, however, was said of any innovations identifying EO as a closed subgroup.

It was Pawley's (1972) paper which made explicit the innovations on which the subgrouping proposal is based, and upon which most subsequent discussion of the subgroup is based.

Pawley (1972) expanded on Biggs' proposal in a number of ways. First, he listed a number of morphological features, and one phonological innovation, apparently shared by the languages of the EO subgroup. Second, he reconstructed

a considerable portion of EO grammar.⁴ Third, he argued for the following subgroups of EO:

- (i) Southeast Solomonian, with two subgroups:
 - a. Guadalcanal-Nggelic
 - b. Cristobal-Malaitan
- (ii) North Hebridean-Central Pacific, with two subgroups:
 - a. North Hebridean, further subdividing into Northern New Hebrides-Banks and Central New Hebrides subgroups;
 - b. Central Pacific, with Fijian and Polynesian subgroups.

Gilbertese and Rotuman were also considered in Pawley's (1972) article, but although both languages satisfied some of the criteria for inclusion within EO, there were some arguments against their inclusion, and they remained unclassified.⁵

Grace, whose 1955 classification of the Oceanic languages was discussed above⁶, reviewed Pawley's (1972) paper in 1976. In this he appeared to accept the main thrust of the argument, although he was critical of a number of aspects of the proposal, in particular the use of the one phonological innovation for the EO subgroup — the loss of Proto-Austronesian (PAN) final consonants in absolute final position. In reaction, Pawley (1977) redefined the subgroup; he withdrew the Southeast Solomonian subgroup, and renamed the rump Remote Oceanic.

More recent work has shown that Pawley may have been somewhat precipitate. Blust (1982), for example, presents a body of evidence suggesting that Pawley's Cristobal-Malaitan subgroup and the Micronesian languages form a subgroup of Eastern Oceanic. While this proposal is inconsistent with Pawley's Southeast Solomons subgroup, it nevertheless suggests that the wider Eastern Oceanic group has considerable validity.⁷

3. CENTRAL-EASTERN OCEANIC

Below we present evidence for a subgroup of Oceanic, called Central-Eastern Oceanic (CEO), comprising the following sets of languages:

1. Eastern Oceanic (EO)
 - a. Southeast Solomonian
 - b. North and Central Vanuatu
 - c. Central Pacific (including Rotuman)
 - d. Micronesian
2. Southern Vanuatu (SV)
3. Utupua and Vanikoro (eastern Outer Islands, Solomon Islands) (UV)

The evidence for CEO consists of a number of morphological features apparently exclusively shared by one or more members of most of the sets of languages shown above, and apparently not found in other Oceanic subgroups.⁸ We use the term 'sets of languages' rather than 'subgroups' because at this stage the internal subgrouping of the proposed CEO group remains uncertain in several respects (although cf. section 4 below).

EASTERN OCEANIC. This set of languages is basically as outlined in the previous section, although given Blust's most recent (1982) paper, the subdivisions of this set listed here must at the moment be taken to be partly geographical and partly genetic.

SOUTHERN VANUATU. This subgroup, which was established by Lynch (1978a), includes the languages of Erromango, Tanna, and Aneityum. Although a certain amount of reconstruction has been done in this subgroup (Lynch 1982c, 1983b), little of this has yet been published.

UTUPUA-VANIKORO. The languages of Utupua and Vanikoro have recently been shown (Tryon and Hackman 1983) to form a rather tenuous subgroup on phonological grounds, although the subgrouping hypothesis is somewhat more secure when morphological evidence is taken into account.

3.1 Exclusively shared features

The following morphological reconstructions are attributed here to PCEO and regarded as innovations of the CEO subgroup:

1. *muni *dative preposition or prepositional verb*
2. *ni *second person singular preverbal subject-marking particle*
3. *le *locative preposition, an irregular development of POC *lo*
4. *(n)tewa *one*
5. *po(η)koto, or *poto(η)ko, *near, nearby*
6. *(q)a *personal article*
7. *ma *preverbal particle marking realis/past/non-future*
8. *tua(i) *verbal particle, 'already, finished, long ago'*
9. *(n)teŋma[na] *particle linking ten and smaller units in numerals above ten*
10. *ŋke *preverbal particle marking conditional*
11. *pia *where? a metathesised form of POC *pai*
12. *tapa *verbal prefix or particle marking spontaneous action*
13. *-(k)i *construct suffix to possessed nouns*
14. *i- *agentive noun derivative*
15. *ka- *instrumental noun derivative*

Of these fifteen reconstructions, the first six or seven seem secure, since they are either found in all the sets of languages noted above, or else their absence from one particular set of languages is explainable in some principled way; in addition, no one set of languages consistently shows more such gaps than any other. The remaining eight or nine shared grammatical features are less secure (and are marked ?PCEO), since they are not found in all sets of languages, although again, there is no one set of languages which consistently misses out on these latter exclusively shared features. We believe that these also provide quite strong evidence for the hypothesis, and may also provide evidence for internal subgrouping at some later date. The exclusively shared features, which are discussed in detail below, are thus ranked from most secure to least secure.

Since there appears to be no significant phonological difference between POC (the ancestor of PCEO) and PEO (a presumed daughter-language of PCEO), we assume for the present that the phonological system of PCEO is the same as that of POC.

The putative innovations are discussed in detail below. Data from each set of languages are presented separately, and this presentation is followed in a number of cases by a discussion of any points of interest which arise.

1. PCEO *muni *dative preposition* or *prepositional verb*

EO: Pawley (1972:8) reconstructed, as an Eastern Oceanic innovation, the form PEO *muni, with the function *prepositional verb indicating dative (to, for, with a person)*, on the basis of reflexes in Fiji (e.g., Wayan muni), Northern Vanuatu and the Banks (e.g., Mota mun, Maewo min) and the South-East Solomons (e.g., Ulawa muni) (1972:88). To the data given there by Pawley can be added the forms Mosina mun, Nume min, and Vatrata men. We have not found cognate forms in Micronesian.⁹

SV: The Tanna languages show morphemes which are formally cognate, although functionally they share only the comitative sense of Pawley's reconstruction. The forms in question are reflected as mone in all five Tanna languages. The Anejom comitative verbal prefix imi- may possibly also be cognate.

UV: The dative forms Vano nimini-, imini-, and Buma nimini- are cognate with *muni.

2. PCEO *ni *second person singular preverbal subject-marking particle*.

EO: The forms Motlav nek, Mota, Vatrata, Mosina, Merlav nik may reflect *ni, though it is more probable that they are reduced forms of the free pronoun (Proto-North Hebridean *(i)-niko (Pawley 1972:113)) rather than true subject-marking particles.

SV: North Tanna, Whitesands, Lenakel, South-West Tanna n-, Anejom nei, na.

UV: Asumboa ni, Tanimbili, Vano nu-, Nembao ni-.

3. PCEO *le *locative preposition*, as an irregular development of POC *lo (cf. PEO *lalo *inside*).

EO: The type PCEO *le appears not to be found in two of the EO regions: South-East Solomons and Central Pacific. It is, however, found in the other two regions. In the Banks, Vatrata, Mosina, Nume, and Merlav all show le. In Nuclear Micronesian, we find Puluwat le or lee *at*, and Woleaian le- *in, prefix to some nouns indicating time or space*.

SV: The locative prepositions in the Southern Vanuatu languages are Sie, Ura ra, North Tanna e, Whitesands, South-West Tanna ie, Lenakel le, Kwamera ia, Anejom a. In each of these languages for which sufficient data are available, a variant form occurs when a pronominal object is suffixed: the forms are Sie, Kwamera, Anejom ira-, Ura ara-, Lenakel la-, South-West Tanna il-. Although there has obviously been some confusion between the palatal (i) and non-palatal (l or r) reflexes of POC liquids in the northern Tanna languages, which appears to have spread to South-West Tanna and Kwamera, it seems nevertheless that these forms are derived from *le.

While the innovation appears to be widespread throughout the putative Central-Eastern Oceanic subgroup, there are also cases of reflexes of POC *lo with a mid back vowel — i.e., cases within CEO languages where the innovation has not taken place. In contrast to the Micronesian forms given earlier, for example, we also find Marshallese ilo. Note in particular the very similar cases of Mota lo, ilo, alo, and Nume le, ile, ale, and also the sets of double reflexes in the same language: Vatrata lo, le; Puluwat llɔ-, lee-, le-. Data from the Utupua and Vanikoro languages are also somewhat confusing. While the

Buma form *lelema* *inside, in the house* is reminiscent of **le*, Asumboa *lu* suggests **lo*, and Nembao *la, ra* suggest **lalo*. A plausible hypothesis is that the change **lo* > **le* was not complete at the time PCEO had broken up, and that the two forms remained variants in at least some dialects of PCEO for some time:

4. PCEO *(n)tewa *one*

EO: Pawley (1972) reconstructed as a North Hebridean-Central Pacific innovation a form for the numeral *one* – variously given as **tu(w)a*, **te(w)a* (p.52) or *(n)tewa (p.113) – on the basis of such forms as East Fijian *dua*, Rotuman *taa*, Merlav *tuwa/le*, Maewo *tewa*, Tangoa *tea*. No cognate forms seem to be found in the South-East Solomons, though there are apparent cognates in Micronesian: cf. Marshallese *juon*, Sonsorol *det*, and possibly also Ulithian *se*, Kusaiean *sie*, *se*, and Woleaian *se-*.

SV: North Tanna, Whitesands *ka/tia*, Kwamera *kwa/tia* may well be cognate.

UV: Nembao *tua*, Tanimbili *suo* are clearly cognate.

Forms reconstructed for *one* include POC *(n)sa and *-kai, and PEO *ke(n)s[ae], *(n)sa, *(n)sakai, and *ta(n)sa. Although a number of these forms are also reflected in various Central-Eastern Oceanic languages, it is plausible that reflexes of more than one such form may occur in the same language with separate functions (e.g., *one, in counting; a, an, indefinite article; one, used in counting tens; etc.*). The presence of reflexes of some of these POC forms for *one* in various CEO languages does not, then, invalidate the proposed innovation.

5. PCEO *(m)po(ŋ)koto or *(m)poto(ŋ)ko, *near, nearby*.

EO: North/Central Vanuatu: Motlav *beteŋ*, Mota *peteŋ*, Vatrata *ro/v'e*, Mosina *ri/pte*, Nume *a/mbəti*, Port Sandwich *paricaŋ*.

SV: Sie *potpot*, Ura *burubut*, North Tanna *i/uakər*, Whitesands *i/uokər*, Lenakel *i/uokət*, South-West Tanna, Kwamera *i/paka*, Anejom *o/potpotet*.

UV: Asumboa *a/vitako*, Tanimbili *ma/voko*, Vano *fatioko*, Buma *fetokɔ*.

6. PCEO *(q)a *personal article*

EO: Pawley (1972:9, 58) reconstructed as an Eastern Oceanic innovation the personal name marker or personal article PEO *(q)a, with reflexes in Polynesian (PPN *(?)a) and the South-East Solomons (e.g., Kwara'ae, Sa'a, Nggela a). A personal article a is also found in Big Nambas in Northern Vanuatu. As a specifically South-East Solomons innovation, Pawley (1972:100) noted that this form occurs only before personal names and kinship terms in focal and subject position (contrasting with the Polynesian function of occurring before personal pronouns, personal interrogatives and personal names in object position).

SV: The Anejom form a, which marks the subject of a clause and is fused initially on the focal pronouns, may parallel the specifically South-East Solomons innovation.

UV: Tanimbili has a personal article of the form a.

7. PCEO *ma *preverbal particle marking realis/past/non-future*

EO: Pawley (1972:48, 112) reconstructed as a North Hebridean-Central Pacific innovation the preverbal particle *ma with reflexes in the Banks and Northern Vanuatu regions (e.g., Merlav *me*, Mota, Nogugu *ma*, Tasiriki *mo*), and also in several Fijian dialects (e.g., Bua, Kadavu *maa*). The form was glossed as marking past or non-future tense. A possible reflex of the form appears to be

found in the South-East Solomons region, in Kwaio *me'e narrative past*. In the Micronesian area, Ulithian *ma habitual* may also be cognate.

SV: All of the Southern Vanuatu languages with the exception of Kwamera show what appears to be a cognate form, though in many cases there is an accreted initial vowel, and in two cases there is a final *n*.¹⁰ The forms in question are Sie, Ura *m- mid past*, *m-em- remote past*; North Tanna *amw-*, Whitesands *am-*, Lenakel *əm-*, South-West Tanna *əmn-*, all marking past tense; and Anejom *mwan perfective*.

UV: No cognate form occurs in the Utupua-Vanikoro languages. However, it appears that these languages have undergone a significant change in their tense-marking system. Essentially, no tense-marking particles may occur in the verb phrase in the languages of Utupua, the time of the action being marked (where necessary) by adverbials:

ASUMBOA: *anambo* *na-mbure-ndyo*
 yesterday I-see-them
 I saw them yesterday

tambo *nene* *na-save* *tavio*
 tomorrow by-and-by I-cut *wood*
 Tomorrow I'll cut the wood

The sole exception to this rule appears to be a particle Nembao *ma-*, Tanimbili *me-*, which marks future. In Vanikoro, on the other hand, tense-marking appears to be effected by a change in the vowel of the subject prefix, especially (though not solely) in the first and second persons singular:

BUMA: *pepane* *ni-romo* *mwoe* *iono*
 yesterday I:PAST-see house your
 I saw your house yesterday

mombo *ne-romo* *mwoe* *iono*
 tomorrow I:FUTURE-see house your
 I will see your house tomorrow

The proposed exclusively shared feature, then, is not strongly contradicted by the Utupua-Vanikoro evidence, which suggests that considerable simplification has taken place in the tense-marking system in these languages.

8. PCEO **tua(i)* verbal particle, 'already, finished, long ago'.

EO: Pawley (1972:9, 51) reconstructed as an Eastern Oceanic innovation the postverbal particle PEO **tua(i)* marking perfective aspect, with reflexes in Polynesian (Proto-Polynesian (PPN) **tuai*), Gilbertese (*tuai*), the South-East Solomons (e.g., Kwara'ae 'ua, Nggela *tua*), and North/Central Vanuatu (e.g., Mota, Tangoa *tuai*). Although the form has apparent cognates in PAN **tuqaS old*, POC, PEO **ma-tuqa old, mature*, and **tuqa-ka older sibling of same sex*, a number of features seem to be innovations: loss of **q*, accretion in at least some languages of *-i*, and its function as an aspect marker. Note also the following forms in Micronesian which are almost certainly cognate in form if not exactly cognate in function: Woleaian *taai no longer*, Sonsorol *mɔ/duwe long ago*. In addition, Pawley established as an innovation of his Central New Hebridean subgroup the change from **tua(i)* to the type Nguna *sua* (1972:119).

SV: Two reflexes of **tua(i)* in Southern Vanuatu are Sie *itetuai*, Anejom *ituwu long ago*. The Sie form shows an accreted *ite-*, which is probably the identificatory prefix *it-* plus epenthetic *e* (Lynch and Capell 1983). The Anejom form is more problematical, and in any case neither is used as an aspect marker per se.

However, the Sie perfective verbal suffix -su may be cognate either with *tua(i) or, more likely, with the Central New Hebridean innovation of the type Nguna sua; in this case, the form is clearly an aspect marker.

UV: Two possible cognates are the perfective aspect markers Asumboa so-, a verbal prefix, and Nembao ao-, a verbal suffix.

9. PCEO *(n)teŋma[na] *particle linking ten and smaller units in numerals above ten.*

EO: Pawley (1972:47) reconstructed the form *(n)tum(w)a, as a numeral particle linking tens and smaller numbers, on the basis of reflexes in Polynesian (PPN *tumaa), Rotuman (fumaa), North/Central Vanuatu (e.g., Merlav demwe/i, Raga domwa) and the South-East Solomons (e.g., Bugotu toma/ga). Additional data from languages not listed in Pawley's tables include Motlav nadmwe, Vatrata wal demwei, Mosina o numweyi, Nume domwen, Big Nambas deman, and possibly Port Sandwich ndromuxer. While Pawley did not establish this form as an Eastern Oceanic innovation, it does not appear to be found outside the putative Central-Eastern Oceanic subgroup.¹¹

UV: Vano lemene, Buma tamana.

The absence of this particle in the languages of Southern Vanuatu can be explained by the fact that these languages have changed to a quinary numeral system. Forms above ten in those languages thus behave no differently from forms above five, and the rationale for a particle linking ten and lower numerals which was distinct from that linking five and lower numerals has disappeared.

10. ?PCEO *ŋke *preverbal particle marking conditional*

EO: The preverbal particle POC *(ŋ)ke, *suppositional, purposive, prospective* was reconstructed as an Eastern Oceanic innovation by Pawley (1972:8-9, 48), on the basis of reflexes in Polynesian (PPN *ke), Fijian (East Fijian, Wayan ke), and the South-East Solomons languages (e.g., Kwara'ae, Lau ke, Nggela ke, ŋge). To these can be added Kusaiean ke *when* (subordinating conjunction), and Gilbertese ŋke, marking conditional.

UV: Asumboa, Tanimbili mbwa-ke *if*; cf. also Tanimbili, Asumboa ke *because*.

It should be noted that this form, either as a preverbal particle or a subordinating conjunction, is represented in all of the regions in question except Southern Vanuatu, where at least the languages of Erromango and Tanna appear to have innovated by using a form of the quotative verb to express conditions. Thus Sie nan̄ku *if*, for example, is a nominalisation of the irrealis form (an̄ku) of the quotative verb oyu; while Lenakel takamwa (real conditional), kapamwa (irreal conditional) are impersonal verbal forms of the quotative verb emwa.

11. ?PCEO *pia *where?* an irregular development involving metathesis of vowels from POC *pai.

EO: Pawley (1972:78, 111-112) posited, as an innovation of the North Hebridean-Central Pacific subgroup, the formal change from POC *pai *where?* to PHC *p[e]ia, based on such reflexes as PPN *fea, Merlav via, Mota vea, and Maewo bea. He notes, however, the difficulty involved in accepting this as an innovation caused by such forms as East Fijian and Wayan vei, which do not show the metathesis. While unmetathesised forms occur in the South-East Solomons (e.g., Arosi hei, Lau fai, Nggela vei), Nuclear Micronesian languages reflect the metathesised form: Gilbertese iia, iaa, Marshallese ia, Puluwat yiya, yiye, Ulithian yiyaa, Woleaian iiya, and Sonsorol iia.

SV: The Southern Vanuatu languages reflect forms with the phoneme combination *ia*, but show irregular loss of the initial **p*. The forms are Sie *iya*, North Tanna, Whitesands *ihia*, Lenakel, South-West Tanna *ihie*, Kwamera *isa*, Anejom *eθa*, suggesting a Proto-Southern Vanuatu form **isia* or **isa*.

Forms recorded in Utupua and Vanikoro are probably more recent innovations, being cognate neither with **pia* nor with **pai* (e.g., Asumboa *ñiñi*, Tanimbili *maa*, Vano *mene*, etc.).

The innovation proposed for PHC thus can be extended to PCEO, with the same kinds of reservations as expressed by Pawley: the problem of the unmetathesised form occurring in the South-East Solomons and Fijian.

12. ?PCEO **tapa* prefix or particle marking spontaneous action.

EO: The forms **ta-*, **tapa-*, *stative derivative*, have been reconstructed for PEO (Pawley 1972:45). A footnote (which refers only to **tapa-* ?) states that 'with some verbs this marks a spontaneously arising condition'. The stative-marker **ta-* is not an EO innovation, as it is found in other Oceanic subgroups.¹² The (variant?) form **tapa-*, however did not appear to be found outside Eastern Oceanic. In Pawley's tables (1972:45) reflexes of **tapa-*, apparently with the meaning of *spontaneity*, are shown as occurring in the Banks and Northern Vanuatu (e.g., Mota *tapa-*, Naewi *rava-*) and also in the South-East Solomons (e.g., Fagani *'afa-*, Bugotu *tava-*). To the data given there can be added Motlav *tav-*, Nume *tava*, Vatrata *'av*, *'amw*, and Mosina *tamw*, *tav*.

SV: Possible cognates of **tapa* occur as suffixes to verbs in the languages of Tanna marking *reflexive*, *reciprocal*, *spontaneously occurring action* — e.g., Lenakel *r-əm-ol-atu* (3SG-PAST-*do*-REC/REF/SPONT) *it happened by itself*. The forms, with accreted initial *a*, are Whitesands *-aru*, Lenakel *-atu*, South-West Tanna *-atukw*, and Kwamera *-atuk*.¹³ Ura of Erromango has the prefix *esebin-* marking spontaneity which may possibly be cognate.

This form does not appear to occur in Micronesian, although the type Marshallese *ja-* may presumably reflect either **tapa-* or the shorter form **ta-*. No affix marking spontaneity has been recorded for the languages of Utupua or Vanikoro and further research is needed to determine (a) whether spontaneity is marked at all in those languages and (b) if it is, whether it is marked at all in those languages and, if so, whether it is marked with a reflex of **tapa-*.

13. ?PCEO *-(*k*)*i* construct suffix to possessed nouns when the possessor is a noun.¹⁴

EO: The following forms appear to derive from the suggested reconstruction. In the South-East Solomons: Kwaio *-i*, *-i*, Nggela *-i* (cf. C.E. Fox 1950:149), and possibly To'ambaita *-e* (Ray 1926). In the North/Central Vanuatu region: Mota *-i*, *-iu*, Vatrata, Mosina, Merlav *-yi*, *-i*. In Central Pacific: East Fijian *-i*.¹⁵ In Nuclear Micronesian: Puluwat *-y*, *-i*, and possibly Ulithian *-li*.

SV: South-West Tanna, Kwamera, and Anejom have a construct suffix *-i*.

Information presently available on the languages of Utupua and Vanikoro shows no evidence of construct suffixes.

14. ?PCEO **i-* agentive noun derivative.

15. ?PCEO **ka-* instrumental noun derivative.

A form **i-* is reconstructed for POC and PEO, and glossed as *instrumental noun derivative*. Within the Eastern Oceanic subgroup, it has reflexes in Fijian,

Northern Vanuatu, and the South-East Solomons. However, there is considerable evidence that a homophonous form *i- can be reconstructed for PCEO with the function *agentive noun derivative*, while a PCEO form *ka- can be reconstructed with the function of *instrumental noun derivative*. The evidence is presented first, and is followed by a discussion of the relationship between these two forms and other forms reconstructed as *i.

Agentive *i:

EO: North/Central Vanuatu: Mota, Motlav, Merlav i, Mosina e.

SV: Whitesands, South-West Tanna, Lenakel, Kwamera i-.

Instrumental *ka:

EO: North/Central Vanuatu: Mota, Mosina, Nume, Merlav ya. Nuclear Micronesian: Sonsorol ya- *forms abstract nouns, including instrumental nouns*. Central Pacific: note Fijian kaa *thing*.

SV: North Tanna, Whitesands, Lenakel, South-West Tanna, Kwamera k-.

It appears that considerable switching of function has occurred in Oceanic languages generally with regard to these two morphemes. The proposed PCEO prefix *ka-, for example, appears to be reflected in some languages with the function of forming agentive rather than instrumental nouns. At the same time, it is our contention that the same kind of reanalysis was taking place with respect to *i, possibly under pressure from a homophonous particle, the personal article PAN, POC, PEO *i. While *i can be reconstructed with purely instrumental functions for PCEO, there is also evidence, as given above, for suggesting that it had acquired, or was acquiring, agentive functions.

Consider the case of Mota (Codrington 1885). It is clear there that both i (from *i-) and a (from *ka-) have instrumental uses. At the same time, there is a certain amount of evidence for the agentive use of i, as in *yale deceive*, *i yale the deceiver*; *ɣilala know*, *i ɣilala the knowing one*; *ɣopa sick*, *yopae sickness*, *i yopae a sick man*; and so on. Similar evidence exists for Motlav and Merlav.

It would seem, therefore, that POC *i was an instrumental prefix. Possibly due to contamination from the homophonous personal article *i, however, this instrumental prefix began taking on agentive-marking functions, as seems to have happened in Mota, and the form *ka- seems to have been innovated to take over the instrumental functions of *i. The process seems to have reached its furthest development in Southern Vanuatu, where *i now has no instrumental uses.

4. CENTRAL-EASTERN OCEANIC: INTERNAL AND EXTERNAL SUBGROUPING

Both Pawley (1977) and Blust (1982) question, in one way or another, the original Eastern Oceanic proposal, and in both cases it is the affiliations of the languages of the South-East Solomons which are the major problem. Pawley proposes withdrawing them altogether from the subgroup, while Blust finds evidence linking Nuclear Micronesian with one of Pawley's (1972) South-East Solomons subgroups, but not with the other. But although the integrity of the Eastern Oceanic subgroup clearly requires considerable further investigation, we have not addressed ourselves to this question in this paper. Rather, in briefly examining the internal subgrouping and the external links of the putative Central-Eastern Oceanic grouping, we assume for the present that Eastern Oceanic is a valid subgroup.¹⁶

An examination of the innovations proposed above suggests two conclusions:

- a. that a strong case exists for including the languages of Southern Vanuatu (SV) in CEO.
- b. that a weaker case exists for including the languages of Utupua and Vanikoro (UV) in CEO.

It is clear that a great deal more work is necessary to clarify the internal subgrouping of Central-Eastern Oceanic.

As for the immediate external relationships of CEO, it is difficult at this stage of the investigation to make any positive statement. However, our impressions — and they are only impressions at this early stage of the investigation — are that the closest relatives of the putative Central-Eastern Oceanic subgroup, an expansion of the original Eastern Oceanic subgroup, will be found on the one hand in the Loyalty Islands and New Caledonia and on the other in the New Ireland/Bougainville/Western Solomons regions, areas geographically contiguous to that covered by Central-Eastern Oceanic.

5. CONCLUSION

We have suggested in this paper that there is a body of evidence supporting the existence of a subgroup which is geographically continuous in central and eastern Oceania. This subgroup clearly includes both the languages generally labelled Eastern Oceanic, and also the Southern Vanuatu subgroup: this conclusion matches exactly that of Grace (1955), though it escaped Pawley (1972) and especially Lynch (1978a).¹⁷ There is also a considerable body of evidence supporting the inclusion of the Utupua-Vanikoro languages in this subgroup.

Obviously, this proposal is a tentative one at this stage, and considerably more research is required to (a) establish the grouping more securely, (b) decide its internal subgrouping, and (c) locate it within the Oceanic family tree.

CENTRAL-EASTERN OCEANIC SOUND CORRESPONDENCES

POC	*p	*mp	*t	*nt	*k	*ŋk
Woleaian	f,∅	p	t,s,∅	ʃ	g,∅	∅
Puluwat	f,∅	p	h,∅	r	k,∅	∅
Sonsorol	f,∅	p	t,d	s	k	∅
Marshallese	y,∅	p,b	j	d	k	k,∅
Kusaiean	∅	p	t,s	sr	k	k
Gilbertese	∅	p	t,∅	r	k,∅	k,∅
Bugotu	v	b	t	t,d	ʎ,g	g
Nggela	v,b	b	t	t,d	ʎ,g	k
Lau	f,b	b	∅	∅,t	ʎ,∅,g	g
Kwara'ae	h,b	b	∅	∅	ʎ,∅,k	k
Kwaio	f,b	b	∅	∅	ʎ,g	g
Ulawa	h,p	p	∅	∅,ts	ʎ,k	
Arosi	h,b	b	∅	∅	ʎ,k	g
Fagani	h,p	p	∅	∅	ʎ,k	k
Nembaō	v,∅		t,r		∅,k	g
Asumboa	v,b		t,s		∅,k	g
Tanimbili	v,∅		t,s	j	∅,k	g
Buma	v,∅,p	b	t,d	t	∅,k	
Vano	v,∅,p	b	t,l	t	∅,k	
Mota	v,w	p	t,s	n(?)	ʎ,w,k	k
Motlav	v,w,p	b	t	d	ʎ,w,k	k
Merlav	v,w	b	t	d	ʎ,w,k	k
Mosina	v,w		t		ʎ,w,k	k
Vatrata	v,w		?	d	ʎ,w,k	k
Nume	v,w	b	t,s	d	ʎ,w,k	k
Cent.Maewo	v	b	t	d	ʎ, v ^w , k	k
Raga	v,p	p	t,s	d	x,k	g
Nogugu	v,w	p,p ^w	t	?	ʎ,k	
Tasiriki	v,w	p	t,k	k	ʎ,k	k
Tangoa	v,ṽ	p,ṽ	t	dr	h,k	k
Big Nambas	v,ṽ	p,ṽ	t,d	d	x,k	k
Pt.Sandwich	v	b	t,dr	dr	ʎ,ŋk	ŋk
Nguna	v,w,p	p	t,d	d	k,ŋ	
Sie	v,p	mp	t,h	t,h	ʎ,k,∅	
Ura	v	b	t,s	s(?)	g,w	
Lenakel	v,w,p	p,p ^w	t,s,r	t	k,∅	k
Whitesands	v,w	p,p ^w	t,s	t,r	k,ŋ,∅	k
SW Tanna	v,w,k ^w	p,p ^w	t,l,s	t	∅,k	k
Kwamera	v,w	p,p ^w	t,r,h	t,r	∅,k	k
N Tanna	v,w,p	mp	t,s	t	k,ŋ,∅	k
Anejom	h,w,p	p,h	t,s	t,ts	x,k,∅	k
Pije	p,v,∅	b,b ^w ,g	t,d,ts	d	ts,k,∅	ts,ŋ
Fwai	p,v,∅	b,b ^w ,g	t,d,ts	d	ts,k,∅	ts,ŋ
Nemi	p,v,∅	b,b ^w ,g	t,d,ts	d	ts,k,∅	ts,ŋ
Jawe	p,v,∅	b,b ^w ,g	ts,j	j	ts,k,∅	ts,ŋ
Rotuman	h	p	f	f	?	k
Fijian	v	b	t,d	d	k	g
Wayan	v	b	t	d	k	g

POC	*pw	*q	*d	*nd	*m	*mw
Woleaian	p	∅	r	ʃ	m	mw
Puluwat	pw	∅	ʃ	r	m	mw
Sonsorol	b	∅	l.r	s	m	mw
Marshallese	b	∅	r	d	m	mw
Kusaiean	f	∅	l	sr	m	m
Gilbertese	pw	∅	∅	r	m	mw
Bugotu	b	∅	r, (d)	r	m	m
Nggela	b	∅, y	r	r	m	m
Lau	gw, b	∅, s	r, d	d	m	ŋ ^w
Kwara'ae	gw	∅, s	r, d	d, r	m	ŋ ^w
Kwaio	g, b	∅, l	r	r	m	w
Ulawá	pw, p	∅, s	r, ts	ts	m	mw
Arosi	bw, b	∅	r, d	d, r	m	mw
Fagani	bw, b	∅	r, t	t, r	m	mw
Nembaó	b	∅	l	l	m	m
Asumboa	p	∅	y, l	l	m	mu
Tanimbili	p	∅	y	l	m	mw
Buma	b, pw	∅	l	m	m	mw
Vano	b, pw	∅	l	m	m	mw
Mota	kpw	∅	r	n(?)	m	mw
Motlav	kw	∅	y	m	m	mw
Merlav	kpw	∅	r	d	m	ŋ ^w
Mosina	kpw	∅	r	m	m	mw
Vatrata	kw	∅	r	d	m	mw
Nume	kw	∅	r	d	m	mw
Cent. Maewo	bw	∅	r	d	m	mw
Raga	pw	∅	r	r	m	mw
Nogugu	p	∅	r	r	m	m
Tasiriki	p	∅	r, k	r	m	m
Tangoa	p	∅	r	r	m, m̄	m, m̄
Big Nambas	p	∅	r, d	r	m, m̄	m, m̄
Pt. Sandwich	b	∅	∅, dr		m	mw
Nguna	pw	∅	r, d		m	mw
Sie	b	∅	r, d	t	m	m
Ura	b	∅	r, t		m	m
Lenakel	p	∅	l, y, t	r	m, mw	mw
Whitesands	p	∅	l, y, r	t	m, mw	mw
SW Tanna	p	∅	r, t	r	m	mw
Kwamera	p	∅	r, t	r	m, mw	mw
N Tanna	p	∅	l, y, t	t	m, mw	mw
Anejom	p	∅	r, ts	r	m, mw	mw
Pije	bw, g	k, ∅	t, l	d	m, mw	m
Fwai	bw, g	k, ∅	t, l	d	m, mw	m
Nemi	bw, g	k, ∅	t, l	d	m, mw	m
Jawe	bw, g	k, ∅	t, l	d	m, mw	m
Rotuman		∅	r	dr	m	
Fijian		∅	r	t, l	m	ŋ
Wayan		∅	r	dr	m	ŋ ^w

POC	*n	*ñ	*ŋ	*s	*ns
Woleaian	l	l	ŋ	t	t
Puluwat	n	n	ŋ	t	t
Sonsorol	n	n	ŋ	t	t
Marshallese	n	n	ŋ	t	t
Kusaiean	n	n	ŋ	t,∅	∅
Gilbertese	n	n	ŋ	r	r
Bugotu	n	ñ	ŋ	∅,h	h
Nggela	n	n	ŋ	h	h
Lau	n	n	ŋ	t,s	t,s
Kwara'ae	n	n	ŋ	t,s	t,s
Kwaio	n	n	ŋ	t,s	t,s
Ulawá	n	n	ŋ	t,s	t,s
Arosi	n	n	ŋ	t,s	t,s
Fagani	n	n	ŋ	t,s	t,s
Nembaó	n	n	n(?)		h
Asumboa	n	n	ŋ	∅	s,∅
Tanimbili	n	n	ŋ	∅	s,∅
Buma	n	n	ŋ	s	
Vano	n		ŋ	s	s
Mota	n	n	ŋ	s	s
Motlav	n	n	ŋ	h	h
Merlav	n	n	ŋ	s	s
Mosina	n	n	ŋ	s	s
Vatrata	n	n	ŋ	s	s
Nume	n	n	ŋ	s	s
Cent.Maewo	n	n	ŋ	s	s
Raga	n	n	ŋ	h	h
Nogugu	n	n	n	s	s,ts
Tasiriki	n	n	n	s	s,ts
Tangoa	n	n	ŋ	s	s,ts
Big Nambas	n	n	n	∅,s	∅,ts
Pt.Sandwich	n	n	ŋ	s	s
Nguna	n	n	ŋ	s	s
Sie	n	n	ŋ	s,(h)	s
Ura	n	n	ŋ	s,(h)	
Lenakel	n,ŋ	n	ŋ,n	s,h	h
Whitesands	n,ŋ	n	ŋ,n	s,h	h
SW Tanna	n,ŋ	n	ŋ,n	s,h	h
Kwamera	n,ŋ	n	ŋ,n	s,h	s,h
N Tanna	n,ŋ	n	ŋ,n	s,h	h
Anejom	n,ñ	n,ñ	ŋ,ñ	s,θ	h,θ
Pije	n	n	n	d	th,t,l
Fwai	n	n	n	d	th,t,l
Nemi	n	n	n	d	th,t,l
Jawe	n	n	n	d	th,t,l
Rotuman	n	n	ŋ	s	s
Fijian	n	n	ŋ	s,ð	ð
Wayan	n	y,∅	ŋ	s	ð

POC	*l	*R	*w	*y
Woleaian	l	Ø, r	w	Ø
Puluwat	l	Ø, r	w	Ø
Sonsorol	n	Ø, l	w	Ø
Marshallese	l	Ø, r	w	Ø
Kusaiean	l	Ø	Ø	Ø
Gilbertese	n	Ø, r	w	Ø
Bugotu	l	l	Ø	Ø
Nggela	l	l	Ø, u	Ø
Lau	l	l	kw	Ø
Kwara'ae	l	l	kw	Ø
Kwaio	r	r	kw	Ø
Ulawa	l	l	w	Ø
Arosi	r	r	w	Ø
Fagani	r	r	w	Ø
Nembaio	l, Ø	y, Ø	v, Ø	Ø
Asumboa	y, Ø	l, Ø	w, Ø	Ø
Tanimbili	l, Ø	y, Ø	w, Ø	Ø
Buna	l, Ø	r, Ø	w, Ø	Ø
Vano	l, Ø	r, Ø	w	Ø
Mota	l	r, Ø	w	Ø
Motlav	l	y	w	Ø
Merlav	l	r, Ø	w	Ø
Mosina	l	r, Ø	w	Ø
Vatrata	l	r, Ø	w	Ø
Nume	l	r, Ø	w	Ø
Cent. Maewo	l	r, Ø	w	y
Raga	l	r, Ø	w	y
Nogugu	l	Ø	w	Ø
Tasiriki	l	Ø	w	Ø
Tangoa	l	Ø	w	
Big Nambas	l	r, Ø	w	Ø
Pt. Sandwich	l	r, Ø	w	Ø
Nguna	l	r, Ø	w	Ø
Sie	l	r, Ø	Ø	
Ura	l			
Lenakel	l, y	l, Ø	w, Ø	
Whitesands	l, y	l, Ø	Ø	
SW Tanna	r	r, Ø	Ø, kw	
Kwamera	r, y	r, Ø	w	
N Tanna	l, y		Ø	
Anejom	l, ts	r, Ø	w	y
Pije	n	Ø	w	
Fwai	n	Ø	w	
Nemi	n	Ø	w	
Jawe	n	Ø	w	
Rotuman	l	Ø	v	
Fijian	l	Ø	w	ö, Ø
Wayan	l	Ø	w	ö, Ø

NOTES

1. Of course a more definitive subgrouping will have to take into account and weigh all of the linguistic evidence, not just morphological innovations. The weighting of this evidence will become crucial at that point, as it will effect the classificatory model adopted ultimately.

Procedures used and languages examined: a checklist of over 230 morphological items or categories has been devised and information will be collected from as many Oceanic languages as possible. To date, fifty-one languages have been investigated in this way, and the data fed into a computer. For the current paper, these fifty-one computer files were compacted, and a preliminary search for possible innovations was conducted. Once the feasibility of the Central-Eastern Oceanic subgroup became apparent, an investigation of possible Central-Eastern Oceanic cognates of the innovations listed in Pawley (1972) was made and also a wider search for innovations in other languages which were potential members of the CEO subgroup was carried out.

The fifty-one languages which have been investigated so far in the manner described above, together with data sources, are listed below: the abbreviation (L) or (T) indicates that data derive respectively from Lynch's and Tryon's fieldnotes:

CENTRAL-EASTERN OCEANIC LANGUAGES.

South-East Solomons: Nggela (Codrington 1885, C.E.Fox 1950); Kwaio (Keesing 1982).

Utupua-Vanikoro: Asumboa (T); Tanimbili (T); Nembao/Amba (T); Vano (T); Buma/Te Ano (T, Peter Lincoln fieldnotes).

North/Central Vanuatu and Banks: Motlav, Mota, Vatrata, Mosina, Nume, Merlav (Codrington 1885); Big Nambas (G.J. Fox 1979); Port Sandwich (Charpentier 1979); Maii (T).

Southern Vanuatu: Sie (Lynch and Capell 1983); Ura (Lynch 1983a); North Tanna (L); Whitesands (L); Lenakel (Lynch 1978b); South-West Tanna (Lynch 1982a); Kwamera (Lindstrom 1982, L); Anejom (Lynch 1982b).

Nuclear Micronesian: Gilbertese (Cowell 1950, Groves, Groves and Jacobs 1985; Marshallese (Zewen 1977); Puluwat (Elbert 1974).

OTHER LANGUAGES.

New Guinea Region: Gitua (Lincoln n.d.); Kaliai-Kove (Counts 1969); Labu (Siegel 1982); Adzera (Holzknecht in press); Central Buang (Hooley 1970); Are (Paisawa, Pagotto and Kale 1976); Dobu (Arnold (1931)); Maisin (Ross 1984, L); Sinagoro (Kolia 1975); Motu (Lister-Turner and Clark n.d., Taylor 1970); Bileki/Nakanai (Johnston 1980); Tangga (Bell 1977); Tigak (Beaumont 1979); Banoni (Lincoln 1976).

Western Solomons Region: Mono-Alu (Boch n.d.); Sengga (T); Roviana (Jones 1949).

New Caledonia: Cemuhi (J-C. Rivierre 1980); Anjie (Lichtenberk 1978); Pije, Fwai, Nemi and Jawe (Haudricourt and Ozanne-Rivierre 1982).

Loyalty Islands: Nengone (Tryon 1967); Iai (Tryon 1968b); Dehu (Tryon 1968a).

Possibly Oceanic languages: Nauruan (Kayser 1938); Yapese (Jensen 1977).

Other potential members of the putative Central-Eastern Oceanic subgroup were then examined in less detail to test whether the presumed Central-Eastern Oceanic innovations were present; in Micronesia - Woleaian (Sohn 1975), and Sonsorol (Capell 1969). More general information was gleaned from Leenhardt (1946), Ray (1926) and Codrington (1885), as well as Pawley (1972).

2. Naturally, these attempts did not involve all the Oceanic languages, but the important feature which distinguishes them from other classifications is that virtually every Oceanic-speaking area is represented.
3. The 1971 and 1981 papers appeared only in mimeographed form.
4. A number of the defining Eastern Oceanic morphological features listed by Pawley can now, with the addition of further data, be shown to be more widespread. Among these are (1) *(w)atu *motion away from speaker*, which has cognates in Roviana atu and Mono-Alu au; (2) *(n)tani *prepositional verb, motion away from object NP*, which is found in Motu -tani; (3) *(C)ana, *(C)a *transforms verb or noun into an abstract noun*, with cognates like Gitua -na, -ana, -zana, Kaliai -na, Labu -ya, and numerous forms of the type Roviana -ana in western Oceanic languages; (4) *ke *demonstrative formative*, with such cognates as Tangga ge *demonstrative, near speaker* and Sengga ye, yi *demonstrative, distant*; and (5) *[ka]Raka *upwards*, with the Banoni cognate yareya. The innovation *ke(n)s(ae) *one, only, apart* may also be invalidated by such forms as Gitua eze, Kaliai ere.
5. Ka'eo (n.d.) has subsequently argued that Nuclear Micronesian qualifies for membership within EO, while the position of Rotuman was clarified by Pawley (1979).
6. Grace's 1955 classification proposed, as one of the nineteen 'subgroups' of Oceanic, a grouping consisting of
 - (i) the languages of Southern Vanuatu;
 - (ii) Fijian, Rotuman, the Polynesian languages, and the languages of Central and North-West Vanuatu;
 - (iii) the languages of North-East Vanuatu and the Banks Islands;
 - (iv) the Nuclear Micronesian languages.
7. Our investigations have suggested that the following six original morphological innovations proposed by Pawley as marking off the EO subgroup are still valid: *su(ldR)i *prepositional verb, 'motion towards or reference to object NP'*; *i *preverbal particle, marking future*; *(n)soko *numeral prefix, 'collectivity, all at once', postnominal particle, 'all, every'*; *-ka, *-a *stative formative*; *-di *demonstrative suffix*; and *kua *today, presently*. In addition, we suggest one further apparent EO morphological innovation, namely ?PEO *(i)ro *feminine (singular) personal article*. Pawley (1972:58, 116) reconstructed for the Northern New Hebrides-Banks subgroup the feminine article *(i)ro, suggesting that while the *i reflects PEO *i, *pronominal article*, the feminine component *-ro is an innovation. To the data given there by Pawley we can add Motlav rV, Vatrata, Nume iro, and Mosina ero. Cognate forms are found, however, within other parts of the EO area: in the South-East Solomons, the Arosi form re- *prefix to names of women* may be cognate; in Nuclear Micronesian, note the forms Marshallese le-, Gilbertese liy-, and Woleaian la-.

8. An explanation of the procedures involved in the comparison, and a list of the languages investigated, together with sources from which the data were drawn, can be found in note 1, above.
9. Pawley (personal communication) indicates that some of the reflexes differ in function from the reconstruction; e.g. Wayan muni is a postverbal or nominal particle meaning *also, as well*.
10. The accreted n may derive from compounding with a past or perfective marker; note Proto-Tanna *(eə)n- *perfective* (Lynch 1982c:17). The presence of the accreted vowel in this form, as in other verbal affixes in these languages, is probably explained by the fact that tense-markers in most of these languages immediately follow person-markers, which generally consist of a single consonant; vowel-epenthesis is common in such cases, and what has been analysed as vowel + m may in fact be epenthetic vowel + m.
11. Wurm and Wilson (1975) give no POC reconstruction although a PAN form *belas is listed (1975:140).
12. Pawley (personal communication) indicates that *ta- also occurs in Indonesian, marking accidental or spontaneously induced states.
13. For the conditioned reflexes Whitesands u, Lenakel u or w, South-West Tanna kw, Kwamera k or kw < POC *p, note the following: POC *ndapu *ashes* > Whitesands nəm/taau, Lenakel nəm/raau, South-West Tanna nəm/lakw, Kwamera nəm/rakw; POC *puaq *fruit* > Whitesands, Lenakel nə/ua-, South-West Tanna, Kwamera nu/kwa-; POC *topu *sugarcane* > Whitesands nə/tu, Lenakel nə/ruw, South-West Tanna nə/tukw, Kwamera nə/ruk (cf. Lynch 1978a). The accreted initial a is discussed in note 10, above.
14. Robin Hooper's paper "Proto-Oceanic *qi" (in this volume) suggests *qi rather than *-(k)i as the reconstruction, and that it may not be confined to CEO.
15. The East Fijian form is found after an inalienably-possessed noun or a possessive article when followed by a proper noun possessor, as in na tama-i Jone *John's father*, na vale ne-i Jone *John's house*.
16. See also note 7, above.
17. Grace (1955) of course excluded the South-East Solomons, see note 6, above.

BIBLIOGRAPHY

ARNOLD, John K.

1931 *A grammar of the language of Edugaura*. Port Moresby: Government Printer.

BEAUMONT, C.H.

1979 *The Tigak language of New Ireland*. PL, B-58.

BELL, F.L.S.

1977 *Tanga-English English-Tanga dictionary*. Oceania Linguistic Monographs 21. Sydney (Introduction by Capell).

- BIGGS, Bruce
 1965 Direct and indirect inheritance in Rotuman. *Lingua* 13:373-405.
- BLUST, Robert A.
 1984 Malaita-Micronesian: an Eastern Oceanic subgroup? *Journal of the Polynesian Society* 93/2:99-140.
- BOCH, Rev. M.
 n.d. A short Alu grammar. Mimeo.
- CAPELL, A.
 1969 *Grammar and vocabulary of the language of Sonsorol-Tobi*. Oceania Linguistic Monographs 12.
 1971 *Arosi Grammar*. PL, B-20.
- CHARPENTIER, J.M.
 1979 *La langue de Port Sandwich*. Paris: SELAF.
- CODRINGTON, R.H.
 1885 *The Melanesian languages*. Oxford: Clarendon Press.
- COUNTS, David T.
 1969 *A grammar of Kaliai-Kove*. Oceanic Linguistic Special Publication No.6.
- COWELL, Reid
 1950 *The structure of Gilbertese*. Beru: Rongorongo Press.
- DYEN, Isidore
 1965 *A lexicostatistical classification of the Austronesian languages*. *International Journal of American Linguistics*, Memoir 19.
- ELBERT, Samuel H.
 1974 *Puluwat grammar*. PL, B-29.
- FOX, C.E.
 1950 Some notes on Nggela grammar. *Journal of the Polynesian Society* 59/2:135-169.
- FOX, G.J.
 1979 *Big Nambas grammar*. PL, B-29.
- GRACE, George W.
 1955 Subgrouping of Malayo-Polynesian: a report of tentative findings. *American Anthropologist* 57:337-339.
 1971 Notes on the phonological history of the Austronesian languages of the Sarmi coast. *Oceanic Linguistics* 10:11-37.
 1976 Review article: Green and Kelly, eds 1972. *Journal of the Polynesian Society* 85/1:103-112.
 1981 Tentative Oceanic subgrouping. Mimeo.
- GREEN, R.C., and Marion KELLY, eds
 1972 *Studies in Oceanic culture history*, vol.3. Pacific Anthropological Records 13. Honolulu: Bernice P. Bishop Museum.

GROVES, Gordon W., Terab'ata GROVES and Roderick JACOBS

1985 *Kiribatese: an outline grammar and vocabulary.* PL, D-64.

HAUDRICOURT, André-Georges

1971 *New Caledonia and the Loyalty Islands.* In Thomas A. Sebeok, ed. *Current trends in linguistics*, vol.8: *Linguistics in Oceania*, 359-396. The Hague: Mouton.

HAUDRICOURT, André-Georges and Françoise Ozanne-Rivierre

1982 *Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie).* Paris: SELAF.

HOLZKNECHT, Susanne

A morphology and grammar of the Amari language, Morobe Province, Papua New Guinea. In press.

HOOLEY, B.A.

1970 *Mapos Buang - Territory of New Guinea.* Ph.D. dissertation, University of Pennsylvania.

JENSEN, John Thayer

1977 *Yapese reference grammar.* Honolulu: University of Hawaii Press.

JOHNSTON, R.L.

1980 *Nakanai of New Britain: the grammar of an Oceanic language.* PL, B-70.

JONES, L.M.

1949 *A Roviana and English dictionary.* Sydney: Epworth.

KA'EO, Karen

1977 *Eastern Oceanic and the Nuclear Micronesian languages.* Typescript.

KAYSER, Fr. W.

1938 *Nauruan grammar.* Mimeo.

KOLIA, J.A.

1975 *A Balawaia grammar sketch and vocabulary.* In T.E. Dutton, ed. *Studies in languages of central and south-east Papua.* PL, C-29: 107:226.

KEESING, R.M.

1985 *Kwaio grammar.* PL, B-88.

LEE, Kee-dong

1975 *Kusaiean reference grammar.* Honolulu: University Press of Hawaii.

LEENHARDT, Maurice

1946 *Langues et dialectes de l'Austro-Mélanésie.* Travaux et Memoirs de l'Institut d'Ethnologie, 46.

LICHTENBERK, Frantisek

1978 *A sketch of Houailou grammar.* *Working Papers in Linguistics*, University of Hawaii 10/2:74-116.

LINCOLN, Peter C.

- 1976 Describing Banoni, an Austronesian language of Southwest Bougainville. Ph.D. dissertation, University of Hawaii.
n.d. Basics of Gitua language. Mimeo, University of Papua New Guinea.

LINDSTROM, Lamont

- 1982 Kwamera (Nyninyfe) dictionary. Computer printout.

LISTER-TURNER, R., and J.B. CLARK

- n.d. *A grammar of the Motu language of Papua* (2nd. edn, edited by Percy Chatterton). Sydney: Government Printer.

LYNCH, John

- 1978a Proto-South Hebridean and Proto-Oceanic. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 717-779. PL, C-61.
1978b *A grammar of Lenakel*. PL, B-55.
1982a. Southwest Tanna grammar outline and vocabulary. PL, A-64:1-91.
1982b Anejom grammar sketch. PL, A-64:93-154.
1982c Some notes on Proto-Tanna morphology. Typescript.
1983a Ura grammar sketch and vocabulary. In Lynch, ed. 1983:148-183.
1983b Preliminary remarks on Proto-Erromangan. In Lynch, ed. 1983:193-220.

LYNCH, John ed.

- 1983 *Studies in the Erromangan languages*. PL, C-79.

LYNCH, John and A. CAPELL

- 1983 Sie grammar outline. In Lynch, ed. 1983:11-74.

MILNER, G.B.

- 1972 *Fijian grammar*. Suva: Government Press.

MOYSE-FAURIE, Claire, and Françoise OZANNE-RIVIERRE

- 1983 Subject case markers and word order in New Caledonia and Loyalty Islands languages. In Amran Halim, Lois Carrington, S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.4: *Thematic variation*, 113-152. PL, C-77.

PAISAWA, Elsie, Louise PAGOTTO and Joan KALE

- 1976 *A short sketch of Are (or Mukawa), Milne Bay, Papua New Guinea*. University of Papua New Guinea Department of Language Occasional Paper No. 2.

PAWLEY, Andrew

- 1972 On the internal relationships of Eastern Oceanic languages. In Green and Kelly, eds 1972:1-142.
1977 On redefining 'Eastern Oceanic'. Mimeo.
1979 *New Evidence on the position of Rotuman*. Working Papers in Anthropology, Archaeology, Linguistics and Maori Studies No. 56. University of Auckland.

RAY, Sidney H.

- 1926 *A comparative study of the Melanesian island languages*. Cambridge: University Press.

RIVIERRE, Jean-Claude

- 1980 *La langue de Touho: phonologie et grammaire du cemuhi (Nouvelle-Calédonie)*. Paris: SELAF.

ROSS, Malcolm

- 1984 Maisin: a preliminary sketch. *PL*, A-69:1-82.

SIEGEL, Jeff

- 1984 Introduction to the Labu language. *PL*, A-69:83-159.

SOHN, Ho-min

- 1975 *Woleaian reference grammar*. Honolulu: University Press of Hawaii.

SOHN, Ho-min and B.W. BENDER

- 1973 *A Ulithian grammar*. *PL*, C-27.

TAYLOR, A.J.

- 1970 Syntax and phonology of Motu (Papua): a transformational approach. Ph.D. thesis, Australian National University.

TRYON, D.T.

- 1967 *Nengone grammar*. *PL*, B-6.
 1968a *Dehu grammar*. *PL*, B-7.
 1968b *Iai grammar*. *PL*, B-8.
 1973 Linguistic subgrouping in the New Hebrides: a preliminary approach. *Oceanic Linguistics* 12:303-351.

TRYON, D.T. and B.D. HACKMAN

- 1983 *Solomon Islands languages: an internal classification*. *PL*, C-72.

WURM, S.A., and B. WILSON

- 1975 *English finderlist of reconstructions in Austronesian languages (post-Brandstetter)*. *PL*, C-33.

ZEWEN, François-Xavier Nicolas

- 1977 *The Marshallese language: a study of its phonology, morphology, and syntax*. Berlin: Dietrich Reimer.

THE STATUS OF PROTO-MICRONESIAN

Byron W. Bender
and
Judith W. Wang

1. INTRODUCTORY¹

Although many of the islands of Micronesia came to the attention of the western world as early as the 16th century, most of the languages of their inhabitants were not discovered by American linguists until after World War II. Work on Micronesian languages has been going on at the University of Hawaii since the 1960s. Much of the earliest effort was directed toward producing language materials for training Peace Corps volunteers for Micronesia (e.g., McCauley 1966, Quackenbush 1966), but some preliminary comparative work was begun in seminars conducted by Byron W. Bender in the late '60s and by George W. Grace in the early '70s. The earlier work concentrated on identifying cognates and plotting their distribution, while Grace's seminars focused on comparing individual Micronesian languages with a higher order proto-language (Proto-Oceanic) reconstructed primarily on the basis of evidence outside Micronesia. Direct comparison between Micronesian languages was begun in 1976-1977 by an informal group of UH faculty and students. This task was greatly facilitated by the use of the computer to compile a unified English to Micronesian 'finderlist' or index for a number of Micronesian language dictionaries that had been processed by computer.

The task of comparison and reconstruction was continued in the summer of 1977 in a Linguistic Institute course conducted by visiting professor Ward Goodenough and in seminars conducted by Byron W. Bender and Robert W. Hsu from 1977 to 1981. Since 1981 there have been no formal meetings of the comparative Micronesian group, but various individuals have continued to work on an independent basis.

A first set of results was presented by Jeffrey C. Marck at the Austronesian Symposium of the 1977 Summer Meeting of the Linguistic Society of America. Marck focused on phonology and lexicon, presenting a preliminary compilation of sound correspondences among the Micronesian languages and a set of tentative reconstructions based on these. A few interesting irregularities were discussed by Marck, but his primary purpose was to describe the regularities that he had observed.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 53-92. *Pacific Linguistics*, C-88, 1985.

© Byron W. Bender and Judith W. Wang

Much of the effort in the later seminars was devoted to entering the accumulated cognate sets into computer storage in a form that can be used with ALIGN, a computer program developed by Robert W. Hsu and James Tharp to extract sound correspondences from the data and display them according to their environment, so that the factors influencing sound change can be more easily determined. To date members of the UH comparative Micronesian project have compiled approximately 1300 cognate sets. Use of the computer as a data-storage device also makes it easier to edit and update the data.

The purpose of this paper is twofold: 1. to describe in relatively non-technical terms the way in which we have used the computer to aid in the task of comparison and reconstruction, and some of the decisions that have been forced upon us in consequence thereof, and 2. to present a preliminary overview of some of the results that we have obtained. Much remains to be done, however, ranging from simple cleanup work on the file to the exploration of both the internal and external relationships of the Micronesian languages.

2. THE DATA

2.1 The languages

The label 'Micronesian' has at least three uses, each of which refers to a somewhat different group of islands and peoples. Geographic Micronesia extends from the former Gilbert Islands (now part of the Republic of Kiribati) in the east past Belau (formerly Palau) in the west to the atoll of Tobi. In between lie Nauru and the Marshall, Caroline, and Marianas island groups.

Political Micronesia refers to the (former) U.N. Trust Territory of the Pacific Islands (TTPI): the Marshall Islands, the various districts of the Federated States of Micronesia (Kosrae, Ponape, Truk, and Yap), the Commonwealth of the Northern Marianas Islands, and the Republic of Belau. Guam, while geographically part of the Marianas island chain, has been administered separately by the U.S. since the Spanish-American War.

In this paper, however, we use the term Micronesian to refer to those languages that Bender (1971) termed 'nuclear', following Matthews (1950) — that is, the languages of geographic Micronesia excluding Chamorro (spoken in Guam and the Northern Marianas Islands), Palauan, and the Polynesian outlier languages of Nukuoro and Kapingamarangi.

Bender labelled two of these languages 'questionably nuclear' — Yapese and Nauruan. Although more is known about these languages now than in 1971, we still are not able to make a definitive statement on the relationship of either to the unquestionably nuclear Micronesian languages. Where available, Yapese and Nauruan forms have been included in the comparative file, but evidence from these languages has not been taken into consideration in the reconstruction of Proto-Micronesian (PMC).

The nuclear Micronesian languages, on which the reconstructions are based, can be divided into five major branches. Three of these consist of single languages: Kiribati (KIR; formerly Gilbertese), Marshallese (MRS), and Kosraean (KSR; formerly Kusaiean). All of these are well represented in the file. The Ponapeic (PP) subgroup includes Ponapean (PNP), Mokilese (MOK), Pingelapese (PNG), and Ngatikese (NGK). PNP and MOK are well represented, while PNG and NGK data are fewer.

The domain of the Trukic (TK) subgroup extends from Truk lagoon to Tobi. E. Quackenbush (1968) estimated there to be at least sixty distinct Trukic speech communities. However, there is the usual question of how many distinct languages are included in this dialect chain. Bender (1971) cites lexicostatistical evidence to show that while the extreme ends of the Trukic continuum are not mutually intelligible and should therefore be considered different languages, it is not so clear where intermediate language boundaries should be drawn. Bender ends up with three Trukic languages: Ulithian (including Sonsorol, Ulithi, and Woleai), Carolinian (including Saipan Carolinian as well as the central Carolinian atolls of Satawal, Pulusuk, Puluwat, Pullap, and Namonuito), and Trukese (the languages/dialects of Truk lagoon, the Mortlocks, and the Hall Islands).

In contrast, Quackenbush (1968) seems to conclude that eleven² languages can be identified (his Fig. 19, pp.106-107) in spite of varying degrees of mutual intelligibility between adjacent languages in the chain. For his study he selected fifteen 'dialect areas' tentatively identified on the basis of available evidence:

1. Sonsorol (representing Sonsorol, Pulo Anna, and Merir)
2. Tobi
3. Falalap, Ulithi (representing Ulithi, Fais, Ngulu, and Sorol)
5. Falalap, Woleai (representing Woleai, Eauripik, Lamotrek, Faraulep, Elato, and Ifaluk)
7. Satawal
8. Saipan (all Saipanese dialects)
9. Puluwat
10. Pulusuk
11. Pullap
12. Ulul, Namonuito
13. Murilo (Hall Islands)
14. Nama (Upper Mortlocks)
15. Moc, Satawan (Lower Mortlocks)
16. Fanapanges (Western Truk)
17. Moen (Eastern Truk)

Mogmog (Ulithi) and Ifaluk were included for 'additional perspective', (p.22) occupying positions 4 and 6, respectively, on the west-to-east list. Saipan Carolinian was given position 8 even though it clearly was the result of emigration from several central Carolinian communities. In spite of an early disclaimer that he will use only the term 'language' "to avoid endless repetition of the phrase 'language or dialect'", Quackenbush's differential treatment of these fifteen dialect areas in his conclusions suggests a position on the dialect language question: Pulusuk (10) is either omitted or hyphenated with Puluwat (9) as "virtually identical"; Saipan (8) is omitted from the idealised map of the areas (cf. Figs. 3 and 4, p.24); Upper and Lower Mortlocks, and Eastern and Western (Lagoon) Truk are each hyphenated as single 'languages' in the chain, in spite of evidence that the members of each pair differ in many features, as would be expected of dialects of the same language. In some ways it appears as though Quackenbush would have preferred to sidestep the language/dialect question altogether.

Lincoln (1981) recognises essentially the same list of eleven languages as did Quackenbush (except for Tobi and Pullap, as distinct from Sonsorol and Puluwat, respectively), while grouping them into Western, Central, and Eastern Trukic in a way that coincides with Bender's Ulithian, Carolinian, and Trukese.

Jackson (1984) assumes there to be at least seven Trukic languages: Lagoon Trukese (TRK), Ulithian (ULI), Pulo Anna (PUA), Mortlockese (MRT), Puluwatese (PUL), Satawalese (STW), and Woleaian (WOL). These plus the two major Saipan Carolinian dialects (CRL and CRN) are relatively well represented in the comparative file. Forms from other TK languages and dialects have been included when available, but no systematic search has been made for such forms.

A fairly careful search has been made for Proto-Oceanic (POC) reconstructions to which the PMC forms may be related. Less effort has been expended in searching for other non-Micronesian forms. We justify this by the fact that our primary goal has been the reconstruction of a plausible ancestral language from which the various Micronesian languages might be descended. Systematic comparisons will have to be made both with its presumed immediate ancestor (POC) and with other presumed daughters of POC in order to determine the exact genetic status and external relationships of Proto-Micronesian.

2.2 Data sources

Dictionaries exist and have been used for KIR, MRS, KSR, MOK, PNP, TRK, PUL, WOL, PUA, and YAP. The Saipan Carolinian dictionary has not yet appeared in print, but the data from the dictionary were available to the project. Other published and unpublished materials have also been consulted. For some languages — PNG, MRT, STW — forms were elicited directly from native speakers.

At this stage in the project most forms in the file have been checked by one or another member of the group. These forms generally have not been coded by source. In many cases this poses no difficulty. For some languages a dictionary or other printed matter is the only source available. When more than one source exists, however, there may be disagreement over form and/or meaning. This is true not only of printed sources, but also of native speakers. Where information about sources has been included it usually takes the form of a person's initials or some similar abbreviation included at the end of a line of data. A partial key has been included at the beginning of the file.

2.3 Representation of forms: computer vs. standard orthography

Ideally, one probably would want to represent all the data in terms of underlying phonological forms. Not all of the languages have been analysed phonologically, however. Among those that have been some use more or less phonemic standard orthographies, while others do not. In at least one case (KSR) the language has resisted several attempts at full analysis.

When the standard orthography is more or less phonemic, we have adopted it — with minor concessions to the limitations of computer processing such as changing characters with diacritics to sequences of characters, e.g., á to A'. Also, because the file originally was entered on punch cards, only upper-case letters were available. Direct communication with the computer via an on-line terminal would permit us to use both upper- and lower-case. Once having chosen to use all upper-case, however, we find it easiest to continue that practice.

In some (but not all) cases where the standard orthography differs not too greatly from the phonemic analysis, we have modified the standard spellings somewhat in the direction of the phonemic representations. Only in the case of

Marshallese, in which the standard spelling differs significantly from the phonemic representations, have we chosen to use the latter instead of the former.

Thus Trukese and Saipan Carolinian are represented in their almost-phonemic standard orthographies.³ Woleaian, too, appears to have an almost-phonemic spelling system, which we have retained in the file.⁴ Elbert's (1972) orthography for Puluwat appears to be phonemic, although he does not say so.⁵ The spelling system proposed by Sohn and Bender (1973) for Ulithi also has a different symbol for each phoneme, but no spelling rules were proposed per se, and this orthography has yet to be used in a dictionary or by Ulithian speakers generally. We have, however, used this system for the ULI data in our file. The orthography used by Oda (1977:Appendix) for Pulo Anna and Sonsorolese is phonemic.⁶

Standard spelling systems do not exist for any of the other Trukic languages or dialects, but all appear to be phonologically similar enough that they can be represented within a single set of general spelling conventions.⁷ The Trukese Orthography Committee specifically chose to adopt an orthographic system that would serve to represent all of the major eastern Trukic languages/dialects (Lagoon Trukese, Mortlockese, the Hall Islands, and Puluwatese), although the Goodenough and Sugita dictionary includes only the central lagoon dialect. We use a modified Trukese orthography for Trukic languages/dialects which do not yet have official spelling systems of their own. Exceptions are made, however, for forms obtained from historical sources such as Lütke.

Among the Ponapeic languages only PNP and MOK have standard orthographies. Rehg and Sohl (1979:xix-xx) indicate that the Ponapean standard orthography is phonemic except that the vowels /e/ and /ɛ/ are both represented by the letter e. These vowels are distinguished in the Northern dialect, but not in Kiti speech; however, information provided by Kenneth Rehg permits us to mark the distinction in our file even though it is not marked in the dictionary. Glides are represented by i and u, resulting in occasional ambiguity (Rehg 1981:50-51).

Harrison (1976:20) indicates that the Mokilese consonant symbols correspond to the consonantal phonemes of the language. As in Ponapean, the vowel phonemes /e/ and /ɛ/ are both represented by the letter e, but are distinguished in the dictionary by the use of different type-faces. The symbol ε not being available on the computer keyboard, we use the ampersand (&) to distinguish /ɛ/ from /e/. Our other major deviation from PNP-MOK standard orthography is in the use of doubled vowels (including doubled digraph oaoa) instead of digraphs with h to represent long vowels. Judging by the information available, PNG and NGK appear to be similar enough to PNP and MOK to be represented within the same overall set of spelling conventions.

Abo et al. (1976) attempted to systematise the various Marshallese spelling practices into a proposed standard orthography for that language. Bender (1968) makes it evident that this orthography is overdifferentiated with respect to the vowels and underdifferentiated with respect to the consonants. Phonemically the language has just three⁸ vowels that differ only in height, whereas the orthography uses nine vowels, showing redundantly the three-way allophonic colouring distinction (front, back unrounded, and back rounded) determined by the surrounding consonants (respectively, plain, velarised, and both velarised and labialised). We employ instead a phonemic transcription based on that given following each headword in Abo et al., which shows the underlying glides - /w/, /y/, and /h/ - that are often omitted from the proposed standard orthography.

Like Marshallese, the Kosraean standard orthography appears to be both over- and underdifferentiated with respect to actual sounds present and to probable underlying forms. The letters and rules for combining them would be almost adequate for broad phonetic transcription if the spelling conventions did not eliminate the distinction among plain, velarised, and labialised consonants everywhere except before the two mid front vowels, orthographic *e* and *ac*. Kosraean has been described as having twelve surface vowels (Lee 1975)⁹:

	front	central unrounded	back rounded
high	i	ih	u
high-mid	e	uc	o
low-mid	ac	uh	oh
low	ah	a	oa

At least one of these (oa) is actually a diphthong which often loses its rounded on-glide in casual speech. Another supposed vowel (uc) appears not to contrast with the lower mid central vowel (uh). However, no one has yet succeeded in producing an adequate phonological analysis of KSR. For lack of a better means of representation, we have adopted the standard orthography for KSR — with some reservations about what it actually may represent.

For Kiribati we have used more or less the spellings given in the Sabatier-Oliva dictionary. Phonologically KIR appears to be less complex than its sister languages, and the orthography reflects this. However, the official spellings given in Sabatier-Oliva oversimplify in certain respects. The official spelling system does not distinguish between plain and velarised bilabial stops (both are spelled with *b*), nor does it mark vowel length, but Sabatier also gave unofficial spellings which provide this information.

The decision to use mostly standard orthographies has resulted in a certain amount of confusion over the phonetic values of various symbols. Both types of difficulties exist in our data: the same symbol may represent different sounds in different languages (e.g., *d* in PNP as opposed to PUA and SNS), or essentially the same sound may be represented by different symbols in different languages (e.g., KSR *sr*, WOL *sh*, and CRL *sch*).

Use of standard orthographies does have the advantage that it makes it easier to locate forms in dictionaries. It is not clear that this outweighs the disadvantage of having to become familiar with a multiplicity of spelling systems. Phonemicisation would help somewhat, inasmuch as it tends to reduce the number of symbols needed for a given language, but even phonemic symbols have some arbitrariness about them — witness the choice of *d* to represent an alveolar fricative in PUA/SNS, or the diacritic " for the combination of velarisation and labialisation in MRS.

3. THE COMPUTER FORMAT: BASIC CONVENTIONS

3.1 Band format for cognate sets

The format and programs that we use are adapted from a more general format and set of programs designed for processing dictionary materials.¹⁰ Each cognate set is treated as though it were a dictionary entry, with the reconstructed PMC form corresponding to the 'headword' of the entry.

Each entry consists of a series of lines of 'bands', each containing one form, normally a putative cognate. Each band consists of three parts: 1. a band label representing the language name (discussed in section 3.2), 2. the putative cognate, and 3. an English gloss. The first band label of an entry (the headword band) is identifiable by the fact that it always begins with a period; headwords of subentries are marked by two periods preceding the band label.

(3.1)	.MC	K	I	N	I	VI. *PINCH, PICK
	KSR	K	IH	N	-	VI. *PINCH, PICK
	TK	K	I	N	I	*PICK UP, *PLUCK (E.G., FRUIT)
	CRL	GH	I	L	I (-)	PICK UP, GATHER
	WOL	G	I	L	I	PICK, PLUCK WITH HANDS
	PUA	K	I	N	I	VI. TO PICK UP (FRUIT)

Thus in the sample entry above the leftmost column, consisting of .MC, KSR, TK, CRL, etc., represents the band labels or language names. The middle column lists the putative cognates, headed by the reconstructed PMC form. The blank spaces which have been inserted between segments serve as boundary markers for the ALIGN and DISPLIGN programs, while the underscores and periods function as placeholders to keep the correspondences aligned (see discussion below). The rightmost column gives the English gloss for each form.¹¹

The data must be entered in such a way that the computer programs will be able to distinguish how the various segments are to be aligned with the corresponding segments in the putative cognate forms. This would be no problem if each segment could be represented by a single character and if all segments in each form corresponded to segments in the cognate forms. Such is not the case, however. Our decision to use the standard orthographies for most languages requires us to make provision for diacritics and digraphs. In any case, not all portions of all forms turn out to be cognate. In reconstructed forms, too — whether our own or those of others — sometimes it is not clear which of two (or occasionally more) proto-phonemes should be reconstructed in a particular instance. There has to be a way to indicate that there are alternate reconstructions for the same segment.

Human judgment currently is required to decide how the various segments correspond among the different forms in a cognate set. In some cases it is not immediately apparent how the segments should be aligned, even when we are reasonably sure that the forms are cognate. By entering the alignment information as part of the data, rather than doing it automatically by program, we can experimentally adjust the alignments on an item-by-item basis in order to improve the overall 'fit' of the correspondence sets.

The actual mechanics of the aligned format are as follows. Segments to be aligned are entered with single blank spaces between them to serve as boundary markers for the ALIGN and DISPLIGN programs. Underscores are used to indicate zero reflexes (presumed loss of a proto-segment). Non-cognate portions of forms are enclosed in parentheses, which the computer program interprets as a signal to 'ignore' those portions in compiling correspondence sets and when displaying the aligned cognate sets. Periods or 'dots' may be used to indicate that a portion of a reconstructed form is missing. All putative cognates within the same set must have the same number of aligned segments (including underscores and dots).

(3.2)	.MC	F	A	K	A	_	A	F	I	*EVENING <*-TIM>
	KSR	_	E	K	UH	_	_	_	_	EVENING
	TK	F	A	K	A	_	_	_	_	*EVENING <*-TIM>
	MRT (LE-)	F	A'	_	_	_	A'	F	_	EVENING
	TRK	F	A'	_	_	_	A'	F	_	EVENING MEAL, MAIN MEAL
	PUL (LE-)	F	A'	_	_	_	_	F	_	EVENING MEAL
	STW	F	A'	_	_	_	_	FF	_	EVENING
	CRL (LEE-)	F	A'	_	_	_	_	FF	_	EVENING
	WOL	F	E	G	A	_	A	F	I	LAST NIGHT
	OC	R.	A	P	I (R.API) EVENING, DUSK

In the above example the underscore in the reconstructed PMC form serves as a place-holder for the *R (represented by R. in our all-upper-case computer key punch orthography) in the presumed POC antecedent *RapiRapi. The initial underscore in the Kosraean (KSR) form represents what appears to be a regular loss of PMC *f in that language. Mortlockese (MRT) and Puluwatese (PUL) le- and Carolinian (CRL) lee- are separate morphemes. CRL lee- occurs in other time words such as leesor *morning*, leealowas *noon*, etc. The use of blank spaces as boundary markers allows us to include the diacritic ' as part of the segment A' (/á/) in various Trukic forms. In these languages /á/ is a low front vowel, distinct from the low central or back vowel /a/. In Satawalese (STW) and CRL the geminate ff corresponds to the single f in other languages and has been aligned as though it were a single segment.

The *_afi part of PMC *faka-afi appears to come from an unreduplicated POC *Rapi, but PMC *faka does not correspond to anything in the presumed POC antecedent. Thus dots are used in the OC band as place-holders for the non-cognate portion of the PMC form, while the reduplicated portion of POC *RapiRapi is enclosed in parentheses to signal that it is to be ignored for alignment purposes.

Alternate reconstructions for a particular position in a proto-form would be given as a single complex 'segment' — that is, delimited by blank spaces, with commas (but no blanks) separating the alternatives. For example,

(3.3) .MC S,S' A K E _ TO *RIDE A VEHICLE

where there is no Kosraean cognate to distinguish between Marck's *s and *S.¹²

Metathesis poses a special problem. That is, in order to show how the segments in a presumably metathesised form correspond with segments in the putative cognates, we have had to undo the metathesis. So as not to lose the actual form, we have enclosed it in parentheses, preceded by a percent sign (%_____), and put it in the gloss portion of the band, e.g.,

(3.4)	.MC	S	U,I	K	U,I	M	A,I	*WRAP, *FOLD
	GIL	R	U	K	U	M	A	FOLD, WRAP UP
	MRS	T	I	K	I	M	_	(%KITIM) WRAP THE BODY; MAT USED TO COVER CORPSES, CASKET
	KSR	SR	O	K	O	M	I	WRAP, ENVELOP (VT)
	PNP	D	I	K	I	M	_	(%KIDIM) WRAP (VT)
	TRK	T	U'	K	U'	M	I	WRAP (VT)
	WOL	T	U'	G	U'	M	I	WRAP (VT)
	PUA	T	U'	K	U'	M	I	WRAP (VT)

For the most part the gloss portion of the band gives only enough information to enable us to identify the particular linguistic form being cited. Specially marked 'keywords' in the gloss portion of the headword band are used as input for compiling an alphabetical English to PMC index — what we call a finderlist. For example, the asterisk preceding 'evening' in the headword (.MC) band in example 3.2 will cause the computer to generate an entry.

(3.5) EVENING

FAKA_AFI <MC>

in the finderlist. The notation '<MC>' following FAKA_AFI indicates that this form has been reconstructed for PMC. We adopted this convention when we began including Proto-Trukic reconstructions not only in the data file, but also in a combined finderlist with the PMC reconstructions.

3.2 Band labels

Band labels consist of 2-4 letters: 2-letter abbreviations for reconstructed proto-languages, 3-letter abbreviations for the names of the various present-day languages, with sometimes an additional character (?, L or X) to distinguish various types of doubtful or superfluous forms which are not used in the reconstructions. (See section 4 for further discussion of the significance of the additional characters.)

Within each entry (cognate set) the languages are entered and displayed in order from east to west, with Jackson's Proto-Trukic (abbreviated as TK) reconstructions immediately preceding the Trukic forms. Nauruan and Yapese forms are listed in positions corresponding to their longitude. Forms from non-Micronesian languages, including reconstructed languages such as Proto-Oceanic, follow in no set order.

The most common bands would be represented in the following order:

- (3.6) .MC = Proto-Micronesian
 GIL = Kiribati (formerly Gilbertese)¹³
 MRS = Marshallese
 KRS = Kosraean (formerly Kusaian)
 PNG = Pingelapese
 MOK = Mokilese
 PNP = Ponapean
 TK = Proto-Trukic
 MRT = Mortlockese
 TRK = lagoon Trukese
 PUL = Puluwatese
 STW = Satawalese
 CRN = northern (Enne; Tanapag) dialect of Saipan Carolinian
 CRL = southern (Elle) dialect of Saipan Carolinian
 WOL = Woleaian
 ULI = Ulithi
 PUA = Pulo Anna
 SNS = Sonsorolese
 YAP = Yapese
 OC = Proto-Oceanic
 EO = Proto-Eastern Oceanic
 FIJ = Fijian

4. FURTHER CONVENTIONS: THE ENCODING OF SUBSIDIARY, NEGATIVE, AND MISSING INFORMATION

4.1 Inflectionally and derivationally related forms - 1

Sometimes a language will contain two or more forms that appear to be reflexes of the same proto-form. For example, many Oceanic languages have a system of 'inalienable' as opposed to 'alienable' possession marking. In the Micronesian languages inalienable possession is marked by putting a personal suffix directly onto the possessed noun. Many of the inalienable nouns also have alienable counterparts that are unsuffixed -- e.g., in order to speak of someone's hand in KSR one would add the appropriate possessive suffix onto a stem *po-* or *paho-*, but to speak about the word for *hand* one would use just *paho*. Sometimes the free forms lack the final stem vowel of the bound forms, sometimes not. Often the vowels differ, not only between free and bound forms, but also among the different possessed forms, according to the (historical) final vowel of the possessive suffix. The synchronic analysis of these possessive forms may be far from straightforward (Rehg 1982). Most analysts would argue that syntactically they constitute two distinct subclasses of nouns. In cases such as these all of the variant (stem) forms may be entered, the one assumed to most closely correspond to the proto-form left unmarked and the other forms marked by an X attached to the language abbreviation in the band label. For example,

(4.1)	.MC	- A - F A R A	*SHOULDER <*-BOD>
	MRS	H A - Y E R A (-)	SHOULDER (COMBINING FORM FOR PERSONAL POSSESSIVE SUFFIXES)
	MRSX	H A - Y E R A (Y)	SHOULDER
	PNP	- A - P A R A (-)	SHOULDER (BASE FORM)
	PNPX	- A - P & R &	SHOULDER (3PS)
	TK	- A U F A R A	*SHOULDER <*-BOD>
	TK	- A - F A R A	*SHOULDER <*-BOD>
	MRT	- A (W) U' F A T A (-N)	SHOULDER (3PS)
	MRTX	- A (W) U' F A R -	SHOULDER
	TRK	- A - F A R A (-N)	SHOULDER (3PS)
	PUL	Y E - F A R A (-N)	SHOULDER (3PS)
	STWX	- E - F A R -	SHOULDER
	STW	- E - F A R A (-N)	SHOULDER (3PS)
	CRL	- A (Y) U' F A R A (-)	SHOULDER
	CRLX	- A I F A R -	SHOULDER
	CRL	- A I F A R A (-L)	SHOULDER (3PS)
	WOL	Y A - F A R A	SHOULDER
	ULI	Y A - F A R A	SHOULDER
	PUA	Y A - D A L A	SHOULDER
	SNS	Y A - F A R A	SHOULDER
	AN	. . . B A R A	SHOULDER
	OC	. . . P A R A	SHOULDER

Note that neither of the variant forms *ayu'fara-* and *aifara-* in CRL is marked by an X. Both will therefore appear in the sound correspondences. Often (but not always) these alternate forms are cross-referenced to each other in the gloss portion of each band, to ensure that the information that an alternate exists is carried along in the concordance data that accompany a compilation of sound correspondences.

The use of the X convention allows us a choice as to whether or not to include these forms when using the ALIGN program to sort out the correspondence sets. For example, the program can be variously instructed to ignore the X bands if we wish to exclude these 'extra' forms, to regard them as a 'different language' from their non-X counterparts if we wish to compare these with the other reflexes, or else to regard them as the 'same language' as the non-X forms if we wish all reflexes to appear together.

In the early stages of the project all forms based on the same verb root were collected together in the same cognate set, with liberal use of the X convention. More recently we chose to separate the various forms into sub-entries according to the process by which they are derived, in the hope that they might be of use in reconstructing part of the morphology of Proto-Micronesian. (See sections 4.6, 7.6.) Removing the X's from the band labels — as we have done — does have the disadvantage of inflating the number of occurrences of those correspondences that are contained within the root. The same correspondences are picked up not only for the root, but for each of the derived forms. This difficulty can be overcome by restoring the X's to all but the first occurrence of the same root in each daughter language.

4.2 Non-cognates

Another use of the X convention is to mark forms that we are reasonably sure are not cognate, but which we wish to include in the file, nonetheless. Many of these are forms which appear to be possible cognates, but which were determined upon closer examination of the correspondences not to be cognate. These have been left in the file to prevent co-workers from re-discovering them and having to re-evaluate their cognacy over and over again. Such forms are not aligned as are true cognates. Instead, the aligned portion of the band is filled by a period or 'dot' in each column, and both form and gloss are given in what would normally be the gloss portion of the band. This prevents them from entering into the sound correspondences even if we choose to include the X bands in a concordance.¹⁴

The same procedure has been used for forms which are clearly not cognate with the rest of a particular set, but which were thought to be of interest for other reasons — e.g., KIR kamea, MRT kamweya, ROT ko'mia *dog* (< Eng. come here ?) as opposed to general Trukic reflexes of PTK *kulaaku (< Cham. gulagu *dog* ?). Most of the clearly non-cognate forms probably should be eliminated from the file eventually.

4.3 Missing or non-cognate portions of forms

Dots have also been used as placeholders in aligned forms which have not been X-ed out, to indicate that only part of the form is cognate. For example Kosraean and Ponapean appear not to have reduplicated the initial CV- of pre-MC *kangi *sharp* as have the other Micronesian languages. This is represented in the following fashion¹⁵:

(4.2)	.MC	K	A	K	A	NG	I	*SHARP
	GIL	K	A	K	A	NG	-	SHARP
	MRS	K	A	K	A	G	-	SHARP
	KSR (LAHL-)	.	.	K	UH	NG	-	SHARP
	PNP	.	.	K	&	NG	-	SHARP
	TK	K	A	K	A	NG	I	*SHARP

To the extent that these dotted portions are coextensive with synchronic or reconstructed morphemes, they reconfirm the morphemic analysis.

4.4 Doubtful cognates

There are some forms whose cognacy is questionable but which were not considered dubious enough to warrant use of the 'X plus dots' convention. These have been aligned, but are marked with a question mark, either at the end of the band label or at the beginning of the gloss.¹⁶ Attaching the question mark to the language abbreviation in the band label permits a choice as to whether or not to include these forms when compiling correspondence sets. Putting the question mark in the gloss portion of the band does not allow such a choice.

4.5 Loanwords

A slightly different problem is the question of how to handle known loanwords. Obvious loans from non-Micronesian languages such as English, Spanish, German, or Japanese generally have been excluded. However, many loans are not at all obvious, particularly when the languages involved are closely related. Where we are reasonably certain that a word is a loan, but want to include it nevertheless, it is marked by an L attached to the band label. For instance, KIR has a rather large number of loanwords from one or more Polynesian languages. Some of these can be identified by the fact that they exhibit a different set of reflexes of certain POC/PMC phonemes. For example, the expected reflex of POC/PMC *r in KIR is \emptyset (loss). In a few forms, however, *r is retained as KIR r. We therefore take many (but not all) of these forms to be Polynesian loans.

In KSR some proto-phonemes appear to be reflected by two or more present-day phonemes, but not enough data exist to establish a pattern of cooccurrences among these multiple reflexes, nor is there any likely donor language in view if we should choose to regard some forms as probable loanwords. It also is possible that the multiple reflexes are a vestige of a former dialect differentiation that no longer exists in KSR. A closer examination of KSR reflexes not only of PMC, but also of POC, may help resolve this question.¹⁷

4.6 Inflectionally and derivationally related forms - 2

The general dictionary format also allows for the use of subentries. Although we have not been entirely consistent in doing so, it is possible to separate out derived forms that are shared by a large number of MC languages, which then can be used to reconstruct a (presumably) derived form in the proto-language. Where enough sets of forms exhibit the same derivational pattern

— that is, where the daughter languages all appear to share the same derivational process — this may suggest the existence of such a process in the proto-language.

In earlier versions of the file all forms with the same verb root were grouped together in a single cognate set with the derived forms X-ed out. All affixes and reduplicated portions were enclosed in parentheses as things to be ignored. Later on it occurred to us that we might be able to reconstruct something beyond the bare verb root if we could compare some of the material that we had been ignoring. In order to do this, these large mixed cognate sets had to be broken up and realigned in smaller sets representing the derivationally related verb classes. However, we still wished to keep such subsets together rather than scattering them throughout the file alphabetically. The subentry convention allows us to do this.

For the main entry we have chosen to use the simplest form of the verb — that is, those forms that we had been reconstructing as the verb roots. These generally are reflected as intransitive verbs in the daughter languages, often less an initial C(V)- that is the result of reduplication. The morphologically more complex forms are treated as subentries. Thus the set formerly represented as

(4.3)	.MC	K	I	N	I		*PINCH, *PICK
	GILX	K	I	N	I	(-KINI)	N. PINCH
	GIL	K	I	N	I	(-KA)	VT. PINCH
	MRS	K	I	N	-	(J-IY)	PINCH WITH FINGERNAILS
	KSR	K	I	H	N	I (S)	VT. PINCH, PICK
	KSRX	K	I	H	N	-	VI. PINCH, PICK
	KSRX	K	I	H	N	(-KIHN)	VI. PINCH, PICK
	MOK	K	I	N	I		VT. PINCH WITH FINGERNAILS
	MORX	K	I	N	I	(-KIN)	PINCH WITH FINGERNAILS
	PNP	K	I	N	I	(-I)	VT. PINCH
	PNPX	K	I	N	I	(-KIN)	PINCH
	TK	K	I	N	I	(-T"-I)	*PLUCK, *HARVEST
	TKX	K	I	N	I	(-KINI)	*PINCH OFF, *BREAK OFF (AS PIECES)

(in which the various Trukic forms are subsumed under the Proto-Trukic reconstructions for the purposes of this example) would be reorganised as a main entry

(4.4)	.MC	K	I	N	I		VI. *PINCH, *PICK
	KSR	K	I	H	N	-	VI. PINCH, PICK
	TK	K	I	N	I		VI. *PICK UP, *PLUCK (E.G., FRUIT)

(plus various Trukic cognates)

with two subentries

(4.5)	..MC	K	I	N	I	K	I	N	I		VI. *PINCH, *PICK	
	GIL	K	I	N	I	K	I	N	-		N. PINCH	
	KSR	K	I	H	N	-	K	I	H	N	-	VI. PINCH, PICK
	MOK	K	I	N	I	K	I	N	-		PINCH WITH FINGERNAILS	
	PNP	K	I	N	I	K	I	N	-		PINCH	
	TK	K	I	N	I	K	I	N	I		ACTION OF CUTTING OR BREAKING OFF PIECES	

(plus various Trukic cognates)

and

(4.6)	..MC	K I N I	-T (-I)	VT. *PINCH, *PICK
	GIL	K I N I	-K (-A)	VT. PINCH
	MRS	K I N	-J (-IY)	PINCH WITH FINGERNAILS
	KSR	K I H N I	-S	VT. PINCH, PICK
	MOK	K I N I	-	VT. PINCH
	PNP	K I N I	- (-I)	VT. PINCH
	TK	K I N I	-T" (-I)	*PLUCK, *HARVEST

(plus various Trukic cognates)

Some of the evidence thus accumulated for various grammatical morphemes in PMC is discussed in section 7.6.

5. COMPUTER-GENERATED APPARATUS: LISTINGS AND SIMPLE SORTS

5.1 DISPLIGN

The DISPLIGN program is designed to produce an easily-read aligned print-out of the data. Thus data that are entered in the form

```
(5.1) .MC M A M A T A *A*WAKE
      GILX . . . . . MAMATA = INTUITION, SAGACIOUS
      MRS M _ M E J _ KEEP AWAKE, STAY UP
      TK M A M A T" A *A*WAKE
      MRTX M _ M A S _ VI. GUARD
      MRT (A-) M _ M A S A' (A'TA') VT. WAKE HIM
      PUL (YA-) . . M A H A (A-LO') VT. TO AWAKEN
      PULX . . M A H A (A-TA') VI. TO WAKE UP
      STW M _ M A S _ AWAKE
      CRLX M _ M A S _ AWAKE
      CRL M _ M A S A (-TA') WAKE UP
      WOL M _ M A T A AWAKE
      PUA M _ M A T A AWAKE
```

with just enough blank spaces to separate the different parts of the band, would appear in the format

```
(5.2) .MC M A M A T A *A*WAKE
      GILX . . . . . MAMATA = INTUITION, SAGACIOUS
      MRS M _ M E J _ KEEP AWAKE, STAY UP
      TK M A M A T" A *A*WAKE
      MRTX M _ M A S _ VI. GUARD
      MRT (A-) M _ M A S A' (A'TA') WAKE HIM
      PUL (YA-) . . M A H A (A-LO') VT. TO AWAKEN
      PULX . . M A H A (A-TA') VI. TO WAKE UP
      STW M _ M A S _ AWAKE
      CRLX M _ M A S _ AWAKE
      CRL M _ M A S A (-TA') WAKE UP
      WOL M _ M A T A AWAKE
      PUA M _ M A T A AWAKE
```

with additional blank spaces inserted for legibility.¹⁸

DISPLIGN also inserts dividing lines between entries, numbers each entry, and counts up the number of aligned segments in each band. When the number of

aligned segments changes within an entry (usually an indication of misalignment and therefore probably incorrect correspondences), this is flagged by the program.

5.2 BANDSORT

If an alphabetical listing of all the forms from a particular daughter language is desired, the BANDSORT program can be used. This program sorts the bands alphabetically, first according to band labels and then by the content of the band, effectively producing a listing in which the data from each language are grouped together in alphabetical order by language name (abbreviation) and, within each language, by form. There is a provision in the program to cause it to select only those bands specified by the user. This kind of printout is useful in proofreading not only in the comparative dataset, but also in ordinary dictionary-making, to ensure correct assignment of band labels and consistency in both labeling and content of bands. An example is given as Appendix A.

5.3 Finderlists

Another program, called INVERT, takes specially marked keywords in the gloss portion of a band and creates an English index or 'finderlist' (see Appendix B). Although the program can be instructed to look in any band or set of bands for its keywords, we have found no particular need for finderlists keyed to anything but the reconstructed Proto-Micronesian and Proto-Trukic forms, since the file is organised according to the reconstructed headwords — .MC in the case of apparently pan-Micronesian cognate sets, .TK in the case of what appear to be exclusively Trukic cognate sets.

A finderlist is not quite the same as the reverse dictionary to be found in many bilingual dictionaries. For one thing, it generally does not give a definition or examples as one would expect in an ordinary dictionary entry, but functions mostly as an index to direct the user to the appropriate entry. More importantly, it may be keyed so as to group together all of the pronouns, for example, or all of the directional suffixes, in addition to listing them according to their nearest English counterparts. It also is possible to code entries so that all forms pertaining to a certain semantic or cultural domain can be listed together under some heading such as geographical terms, parts of the body, names of plants, etc.

6. COMPUTER-GENERATED APPARATUS: CONCORDANCES OF CORRESPONDENCES

As anyone knows who has ever had to compile correspondences by hand, this is a laborious, monotonous, and time-consuming chore — precisely the sort of task one wishes to relegate to a computer. It was for this purpose that the ALIGN program was conceived.

The principle according to which the ALIGN program compiles sets of sound correspondences is the same as is used to produce text concordances. That is, one must locate all occurrences of the item in question, determine what constitutes a relevant context or environment, and sort the various occurrences

of that item according to contexts. In doing ordinary text concordances this means listing all the phrases or sentences in which the desired word occurs. In doing historical-comparative phonology this means listing all the cognate sets in which the desired correspondence occurs.

The ALIGN program recognises as a 'segment' any character or sequence of characters delimited by single blank spaces on either side. Thus, long vowels and geminate consonants may be regarded as different from their short counterparts — e.g., X YY Z vs. X Y Z — or one may choose to represent the long/geminate segments as sequences of two identical short segments — X Y Y Z. This format also accommodates (other) digraphs and diacritics — when these are represented as sequences of characters (e.g., A') — as well as alternate segments (e.g., .MC A S,S' A *rub* — cf. POC *asa *grate*, *sharpen by grating or rubbing*). Parentheses may be used to enclose non-cognate portions of forms; such material will be 'ignored' — i.e., omitted from the correspondences.

The output of the ALIGN program consists of an index of the correspondence sets, plus a concordance or listing of the forms grouped according to correspondences. Within each group of forms the segments being concorded — that is, the correspondences being displayed — are vertically aligned, and the forms ordered alphabetically according to following segments.¹⁹ This permits the reader to examine all occurrences of a correspondence in a given environment without having to search through the entire file for the forms. A sample concordance page is given as Appendix C, and a few excerpts are given in section 7.4.

A concordance may involve any number of languages, in any order. The number of formally different correspondences increases appreciably with the number of languages being concorded. This is partly explained by the fact that the ALIGN program treats a correspondence with a gap (one that is lacking a cognate in a particular language) as something different from a correspondence which is identical except that the gap is filled (one for which all languages have a putative cognate), and partly by the fact that the different reflexes of a proto-segment in one daughter language will not, in general, be divided among the etyma in the same way as the reflexes in another daughter language (one instance of this is discussed in section 7.5). Some of the increase in the number of correspondences is simply 'noise' — one or two occurrences of an irregular correspondence which may represent anything from a keypunch error or error in alignment to a loanword whose non-native origin is revealed thereby.

7. THE RECONSTRUCTIONS

7.1 Marck (1977)

Marck (1977) includes approximately 300 tentative reconstructions based on the following correspondences:

PMC	*p	*p'	*m	*m'	*f	*t	*t'	*s	*S
GIL	p	p'	m	m'	∅	t,∅ ²	r	r	r
MAR	p	p'	m	m'	y	j	d	t	t
KUS	p	f	m,∅	m	∅	t,s ³	ʒ	t	∅
MOK	p	p'	m	m'	p ¹ ,∅	j,∅ ⁴	s	t	t
PON	p	p'	m	m'	p ¹ ,∅	s,∅ ⁴	t'	t	t
MUR	p	p'	m	m'	f	s,∅ ⁵	r',cc	t	t
TRK	p	p'	m	m'	f	s,∅ ⁵	c	t	t

PUL	p	p'	m	m'	f	h, \emptyset^5	r', cc	t	t
CAR	p	b', pp'	m	m'	f	s, \emptyset^5	ʂ, cc	t	t
WOL	p	b', pp'	m	m'	f	t, s ⁵	ʂ, cc	t	t
UTH	p	b', pp'	m	m'	f	t, s ⁵	c	θ	θ
SON	p	b', pp'	m	m'	f	t, \emptyset^5	s, cc	t	t
PMC	*l	*n	*r	*k	*x	*ŋ	*ñ		
GIL	n	n	∅	k, \emptyset^2	∅	ŋ	n		
MRS	l, l	n, ʌ^1	r	k	∅	ŋ	n		
KUS	l	n	l, r ⁸	k	k	ŋ	∅, l ¹		
MOK	l	n	r	k	r	ŋ	∅		
PON	l	n	r	k	r	ŋ	∅		
MUR	l	n	r	k, s ⁶ , \emptyset^1	∅	ŋ	n		
TRK	n	n	r	k, s ⁶ , \emptyset^1	∅	ŋ, n ⁷	n		
PUL	l	n	r	k, \emptyset^1	∅	ŋ	n		
CAR	l	l	r	g, kk	∅	ŋ	n		
WOL	l, nn	l, nn	r	g, kk	∅	ŋ	n		
UTH	l	r, l	r	g, kk	∅	ŋ	n		
SON	l	r	l	g, kk	∅	ŋ	n		

¹ /___ a

² See Marck's discussion (p.17) of *t and *k deletion in Kiribati.

³ /___ (i and a)

⁴ /___ (i, u, and e)

⁵ /___ (non-low vowels)

⁶ /___ i (occasionally)

⁷ /___ i

⁸ less prominent than l but not rare.

Most of these correspondences are reasonably well-attested. Marck did, however, reconstruct two proto-phonemes on the basis of relatively few reflexes. *x was reconstructed solely on the basis of reflexes of *waka *canoe*, but is supported by the presence of prenasalised *-ŋk in Proto-Oceanic. The first person singular possessive suffix was subsequently reconstructed as *-xi (cf. POC *-ŋku).

Marck also reconstructed a distinction between PMC *s and *ʂ on the basis of Kosraean, in which *ʂ has been lost. This he justified by the apparent correlation between KSR loss and POC *ns. It now appears that the correlation is not as straightforward as Marck had thought (see section 7.5). Part of the difficulty is due to disagreement among the various Oceanic languages. Prenasalisation is problematic in Oceanic, in any case. The POC *s/*ns distinction is even more problematic because there is evidence to suggest that what has been reconstructed as *ns may, in fact, not be simply the prenasalised counterpart of *s (Ward Goodenough, p.c.).

Another problematic correspondence set is one which Marck thought might turn out to represent a distinctively Micronesian third palatal, different from that reconstructed by Blust (1976) for Proto-Fijian-Polynesian and apparently also different from Milke's POC *nj. More data are available to us now which suggest that Marck may have been on the right track (see section 7.4). Marck also found some evidence for a rounded k (*k') in Proto-Micronesian.

For the vowels Marck reconstructed the same five that others have reconstructed for Proto-Oceanic: *i, *e, *a, *o, *u — with back and central allophones of the last conditioned by the preceding sound. Preceding *p', *m', *t', *S, *k, *r, *ŋ, *w, *o, and *u allowed *u to remain back, while preceding *t, *s, *l, *n, *ñ, *i, and *e caused *u to be centralised and possibly unrounded. The labials *p, *m, and *f apparently did not occur before round vowels. Other consonants such as *s, *k', and the putative third palatal were left unclassified due to lack of relevant data. Many of the daughter languages have a high central vowel which must be regarded as a separate phoneme synchronically.

7.2 Reconstructions since Marck (1977)

Marck's reconstructions and the cognate sets on which they were based form the core of the computerised comparative Micronesian file. As the file expanded additional reconstructions by Jackson, Trussel, and Wang (among others) have been based on the correspondences compiled by Marck.

The file has grown appreciably since 1977. It now consists of approximately 1300 cognate sets with considerably more supporting data than were available to Marck. With the help of the ALIGN program it is possible to re-examine the correspondences observed by Marck and perhaps to discover some that he missed. As a result it may prove necessary to revise a number of reconstructions. A clearer picture of the ancestral language should result from this exercise.

The first step in this re-examination consists of a careful scrutiny of the correspondence sets compiled by the ALIGN program.

7.3 NEWMIC4A concordances: procedure

Following Bender (1971), we chose to concord on the five presumed major branches of the Micronesian language family: Trukic, Ponapeic, Marshallese, Kiribati, and Kosraean. Trukic is represented by Jackson's Proto-Trukic reconstructions. Ponapeic is represented by Ponapean in the absence of a reconstructed Proto-Ponapeic. The other branches consist of a single language each.

Dyen's (1965) 36.1% cognate percentage between Trukese and Ponapean suggests a close relationship between these two languages, so we expected to find greater agreement between them than among the other languages. A two-way TK-PNP concordance produced approximately 200 formally or mechanically different correspondences.

Marshallese was the next language to be included in the concordances. MRS has merged the proto-vowels to four phonemes which differ only in height, their colour (front-/backness and rounding) being determined synchronically by the surrounding consonants. The consonants, on the other hand, had multiplied themselves through splits as a result of the 'reading off' of colour from the proto-vowels. There are now three distinct series of consonants in MRS: plain but phonetically palatalised, velarised, and both velarised and labialised (see Bender 1968 for details). These historical changes are reflected by an increase in the number of different correspondences to approximately 500 in the three-way concordance.

Kiribati has undergone a number of mergers among the consonants, but it appears to have preserved an earlier vowel system, as may be seen if one compares present-day KIR forms with the forms reconstructed for Proto-Oceanic. Nonetheless, adding KIR to the concordance caused the number of separate correspondences to increase to over 1300.

We chose to put Kosraean last because of our suspicion of the existence of multiple reflexes in that language (even more than are to be found in KIR as a result of Polynesian borrowings). Adding KSR increased the number of different correspondences to over 2500. If KSR had been put first the irregularities in this language would have scattered some of the regularities among the other languages. The total number of distinct correspondences would have remained the same, but it would have been more difficult to perceive areas of agreement among the four other languages if these were split among a number of different reflexes in KSR. (One instance of this is described in section 7.5).

There is inevitably a certain amount of apparent conflict in these results. Part of this is systematic variation among alternate reflexes. Examination of the forms in the concordance data may reveal whether this is systematic (conditioned) or not. Lack of agreement in the environments in which alternate reflexes occur may be interpreted in a number of different ways — unconditioned split is one possible explanation. Dialect differences, or other variation in the daughter language are other possible explanations.

Another kind of apparent conflict is caused by missing cognates where a form has been lost or simply not discovered in a daughter language. We make the assumption that unless these incomplete correspondence sets conflict with the full correspondences in which all languages being concorded are represented, they should be considered to agree with the latter and can be incorporated into them, thus increasing the number of examples that can be referred to in attempting to discover conditioning factors for sound changes.

The real 'noise' — typographical errors, misalignments, highly dubious cognacy, etc. — usually manifests itself in the form of one or two occurrences of a relatively improbable-looking correspondence. These we have simply ignored.

To arrive at the results presented in 7.4 we began with the four-way concordance of sound correspondences involving Proto-Trukic, Ponapean, Marshallese, and Kiribati. This included some 1300 formally different correspondences.

From these we selected those consonant correspondences for which all four languages showed a cognate and which occurred at least three times.²⁰ Correspondences which occurred only once or twice were dismissed as probably 'noise'.

To our selected correspondences we then added those from which one or two cognates were lacking — provided that the remaining reflexes did not conflict with a previously established correspondence. In order to examine all of the environments in which these correspondences occurred, we had to re-group and re-order the relevant portions of the concordance data. A special program called REGROUP was written for this purpose.

From our examination of the environments we concluded that some correspondences could be merged, while others should remain apart for the time being. The correspondences remaining after all plausible mergers were made have been labelled with proto-phonemes. In some cases subnumbers have been assigned to different subcorrespondences within a proto-phoneme — e.g., *s, *s₁, *s₂ (discussed in section 7.4).

The next step was to add in the fifth major branch, Kosraean. The five-way concordance included approximately 2500 formally different correspondences. Again we began with those correspondences involving all five languages and occurring at least three times. In addition, some full correspondences which occurred only once or twice were included on the basis of their greater frequency in the four-way concordance. Non-conflicting correspondences with only one or two gaps were added in cases where Kosraean showed a different distribution of reflexes from the other languages. In cases where Kosraean had a single reflex in agreement with the other languages additional data were not sought out. A portion of the results from the five-way concordance is sketched out in section 7.5.

7.4 NEWMIC4A concordances: preliminary results from the four-way correspondences

We present here those consonant correspondences among the four languages (Proto-) Trukic, Ponapean, Marshallese, and Kiribati that recurred at least three times and for which each of the four languages had cognates. The quantities given in parentheses following each correspondence are for such full correspondences. Also available to us for detailed study of a given correspondence are the instances that show one or two gaps, that is, correspondences in which only three of the four languages, or two of the four, showed cognates. The proto-phonemes used to label each correspondence are generally those used by Marck (1977), although the subnumbers have been added by us. The left-to-right order of the languages is TK-PNP-MRS-KIR.

Labial Obstruents

*p	p/p/p/b	(25)
*p'	p'/pw/b/b'	(9)
	p'/pw/b/b	(10)
*f	f/p/y/∅	(10)

The second *p' correspondence is primarily an artifact of the neutralisation of the b/b' contrast in Kiribati before u and o – the velarised variety occurs in such environments but is unmarked in the transcription. However, there are eight occurrences of this correspondence with non-round vowels for which the KIR facts need further checking.

Apical Obstruents

*t'	c/t/d/r	(21)
*s	t/d/t/r	(47)
*s ₁	t/d/j/r	(3)
*s ₂	t/d/y/r	(3)
*t	t"/s/j/t	(34)
*t ₁	t"/ ₁ /s/j/t	(4)
*t ₂	t"/∅/j/t	(15)
*t ₃	t"/./t/. ²¹	(5)
*t ₄	t"/./y/.	(3)

The *t' correspondence is relatively stable in comparison with *s and *t, which proliferate into seemingly related but deviant correspondences for which conditioning factors have yet to be found, or operate sporadically, and to which we assign subnumbers for the purposes of this discussion.

*t₂ results from the tendency in Ponapean to delete s before high vowels. It has not run its course through the lexicon, as PNP suk *to pound*, sing *fart*, etc. attest, while at the same time it has begun to affect s's before mid vowels, as for example in mɛɛ(-la) *to die* < PMC *mate (cf. TK *mat'e *dead, die*, MRS me'j *dead*, KIR mate *dead, death*). In cases of this sort, which seem to need the concept of lexical diffusion (Wang 1979) to explain contrasting reflexes in a given environment, subnumbers can serve to record the progress of a tendency in each etymon.

The same tendency has progressed further in the eastern Trukic languages, so that the expectation is that s will be lost (or replaced by a glide) before all non-low vowels, as in TRK máá(-nó) *to die* (cf. WOL mase *to die*), although here again it is still to be found in some words, e.g., TRK so *alight* < PTK *t₁oko (cf. PNP sok, MRS jok *alight*, KIR toka *be placed on*). As this example shows, we use the subnumber one to label those PTK *t's that are exceptional in having s reflexes before non-low vowels in eastern TK languages such as Lagoon Trukese and Mortlockese. This in turn accounts for the existence of PMC *t₁.

The present treatment of Trukic and Ponapeic is obviously not parallel. Eventually we would hope to reconstruct a Proto-Ponapeic and to confine to those reconstructions — at least initially — the proliferations resulting from internal Ponapeic developments. Such parallel treatment would reveal more clearly the progress of lexical diffusion in each of the two subgroups, and whether or not given etyma were affected in parallel fashion.

*s₁ and *s₂ may result from the PAN palatals having had more than one reflex in PMC, or other factors may have been involved. The three full correspondences of *s₁ (e.g., TK *p'otau *small basket*, PNP ohdow *basket*, MRS bejaw *pocket, pouch, hand basket of small weave*, KIR b'ara *small cap-like basket made of coconut leaves*) might seem little more than sporadic exceptions, were they not augmented by others having gaps, as shown in the following excerpts from the four-language concordance²²:

(7.1)	t/.j/.	(5)		
	FARM, CLEARING	TK	M A _ A T A	
	NO COGNATE	PNP		
	CLEARED SPACE, OPEN FIELD	MRS	M A H A J _	
	NO COGNATE	GIL		
	VT. HELP	TK	T A P A -NG (-I-)	
	NO COGNATE	PNP		
	VT. HELP SOMEONE	MRS	J I P A -G (-EY)	
	NO COGNATE	GIL		
	SHINE, LIGHT	TK	T I N A _	
	NO COGNATE	PNP		
	START A FIRE	MRS	J E' N _ _	
	NO COGNATE	GIL		
	STING, SMART	TK	T O N G O	
	NO COGNATE	PNP		
	PAIN IN ARM...	MRS	J E G" _	
	NO COGNATE	GIL		
	INSIDE, CORE, INTERIOR	TK	_ U T U	
	NO COGNATE	PNP		
	INTERIOR OF AN ISLAND	MRS	(YA-) W A J _	
	NO COGNATE	GIL		

(7.2) t/d/j/. (2)

BECOME SHALLOW	TK	P E T E P E T E
SHALLOW	PNP	P & D & P & D -
SHALLOW, SUPERFICIAL	MRS	P E' J - P E' J -
NO COGNATE	GIL	

BECOME SHALLOW	TK	P E T E P E T E
SHALLOW	PNP	P & D & P & D -
SHALLOW, SUPERFICIAL	MRS	P E' J - P E' J -
NO COGNATE	GIL	

MORAY EEL	TK	L A P' U T O
SALTWATER EEL (GENERIC)	PNP	L A P W & D -
MORAY EEL	MRS (KIDE'DDE'L)	- - B I J -
RABONO = EEL	GIL

(7.3) ./d/j/. (1)

NO COGNATE	TK	
TO SAVE FROM HARM	PNP	D O A R E
REANIMATE, RESTORE, REVIVE	MRS	J A R I (Q)
NO COGNATE	GIL	

(7.4) ././j/r (1)

NO COGNATE	TK	
NO COGNATE	PNP	
HUSK A COCONUT W/ TEETH	MRS	H E J E -K
LOP BRANCHES, TRIM, PRUNE	GIL	- E R E -

We see here one of the many advantages of computerisation for tasks such as these: the ability to organise and gain ready access to large sets of complex data.²³

*t₃ and *t₄ further exemplify this point. Although attested by no full correspondences, they were brought to light as deviations from the *t correspondence by holding constant the TK and MRS reflexes while permitting gaps in either PNP or KIR or both. Five instances of the former and three of the latter were thereby unearthed: e.g., TK *t"aro- *near*, MRS tir"i- *near*; TK *kut"u' *louse*, MRS (ya-)kit *delouse* (cf. MRS kij *louse* in the regular *t correspondence): TK *ku'u't"a, PNP kiis, MRS qe'ye't *octopus*; TK *t"aru *oyster*, *oyster shell*, MRS yar" *oyster*; TK *pet"i *float*, MRS pey *drift*, KIR beibeti *float*. The explanations for these MRS deviations, like those in *s₁ and *s₂, remain to be uncovered.

Although the PAN palatals generally are reflected as *s, several instances of their being reflected as *t (or one of its subcorrespondences) have appeared in addition to PMC *t₁iri *to spurt, urine* (< PAN *cirit *spray out*) noted by Dyen (1949) and discussed by Goodenough (1961) and Blust (1978). These include PMC *t₁up'e *catch* (TK *t"up'e(-li) *catch*, MRS jibe(-y) *hold, grasp, seize, capture*) < PAN *zambat *carry, hold*, and PMC *tuli *shoot or sucker from a root crop* (TK *t"ili *sprout, shoot from a root crop*, PNP ili *sucker of a banana, breadfruit, taro, etc.*, MRS jil" *shoot, bud, sprout*) < PAN *suli *sucker*. For the latter MRS presents a doublet in the regular *t correspondence: yil *taro sprout*. PMC *t₁iri is attested in three of the four languages: TK *t"iri

urinate, masturbate, MRS (j-)jir *slippery, lubrication*, KIR tii *to spurt, to spout*, and has the doublet PMC *t₁iri *masturbate*: TK *t''iri(-i) *masturbate*, MRS yiri(-y) *wipe* (cf. KSR iri *masturbate*).

Velar Obstruents

*k	k/k/k/k	(66)
	k/k/k/∅	(8)
*k ₁	k/k/q/k	(12)

The second *k correspondence results from the tendency in KIR to delete k or t to prevent their cooccurrence in the same root (or more than one instance of either in a root) — seemingly a sporadic matter to which there are exceptions (Marck 1977, Trussel n.d.). *k₁ shows considerable complementarity with *k, the former generally appearing before round vowels, but there are sufficient counterexamples to warrant caution in merging them. For example, only KSR koet shows evidence of former vowel rounding in the word for 'octopus' (cited above) in which MRS shows q, and both *k and *k₁ are found before *u.

Nasals

*m	m/m/m/m	(43)
*m'	m'/mw/m'/m'	(7)
	m'/mw/m'/m	(8)
*n	n/n/n/n	(22)
	n/n/n'/n	(0) ²⁴
	n/n/n''/n	(1)
*ŋ	ŋ/ŋ/g/ŋ	(19)
	ŋ/ŋ/g''/ŋ	(8)

The situation for the *m' correspondences parallels that for *p', except that there are fewer counterexamples. The second and third *n correspondences result from the read-off of vowel colour onto consonants in the development of Marshallese, and can probably be merged as part of a fairly elaborate scenario that has been proposed by Kenneth Rehg (UH seminar presentation, 1979). As nearly as can be determined, the two *ŋ correspondences seem more capable of being merged than do *k and *k₁.

Liquids

*l	l/l/l/n	(33)
	l/l/l'/n	(19)
	l/l/l''/n	(0) ²⁴
*r	r/r/r/∅	(22)
	r/r/r''/∅	(4)
*r ₁	r/r/r/r	(3)

The *l correspondences parallel the *n correspondences in every respect. The two *r correspondences seem capable of merger in a way paralleling *ŋ, based on the complementarity of rounding in neighbouring vowels. *r₁ marks the seemingly sporadic retention of r in KIR. The three full correspondences are augmented by five others with one gap each, e.g., *p'aro *box, container* (TK *p'aro *box, crate, strong container*, PNP pwoahr *hole, cave*, KIR baro *provision box, overflow, hole, depression where water stays* (cf. baron te wa *forepart of the canoe which dips up water in rough seas*)).

Glides and Zero

*w	w/w/w/w	(8)
	w/w/w/∅	(7)
	w/∅/w/∅	(9)
	∅/w/w/∅	(7)
*∅	∅/∅/∅/∅	(27)
	∅/∅/ɣ/∅	(45)
	∅/∅/ħ/∅	(26)

These correspondences shade from the *w of considerable antiquity in etyma such as PMC *wakara *root* (< POC *wakaRa) and *maSawa *ocean* (< POC *masawa), to what are clearly prothetic or epenthetic glides in Marshallese alone, which has gone farthest among the nuclear Micronesian languages in reading off vowel colour onto adjacent consonants. Much remains to be done in Austronesian generally to settle questions having to do with the status of onsets and glides. Further progress for Micronesian than what we sketch here must await resolution of the many questions regarding the vowel correspondences that still confront us.

7.5 NEWMIC4A concordances: five-way correspondences

The most significant effect of adding Kosraean to the concordance was to increase the number of different correspondences. Some of this was just 'noise', but there were also a number of correspondences that met our criteria (completeness and frequency) for further consideration. In several instances Kosraean split (or further split) proto-phonemes established on the basis of the four-way correspondences. Only one such instance will be presented here.

In section 7.4 we tentatively reconstructed PMC *s on the basis of the correspondence (TK/PNP/MRS/KIR) t/d/t/r. Two subcorrespondences (t/d/j/r and t/d/y/r) were labelled *s₁ and *s₂, respectively. Marck (1977) had recognised two correspondences (TRK/PNP/MRS/KIR/KSR) t/d/t/r/t and t/d/t/r/∅ which he labelled *s and *S, respectively. The five-way concordance yielded the following correspondences:

PMC	TK	PNP	MRS	KIR	KSR	N ²⁵
*s	t	d	t	r	t	19
	t	d	t	r	∅	17
	t	d	t	r	s	12
	t	d	t	r	sr	8
	t	d	t	r	y	3
*s ₁	t	d	j	r	t	3
	t	d	j	r	s	2
	t	d	j	r	sr	1
*s ₂	t	d	y	r	∅	2
	t	d	y	r	sr	1

If we ignore the different MRS reflexes and recast the distribution in terms of the KSR reflexes we still encounter certain difficulties:

*s (?)	t	d	r	t	22
*s (?)	t	d	r	∅	19

* ?	t	d	r	s	14
* ?	t	d	r	sr	10
* ?	t	d	r	y	3

Both MRS and KSR show evidence of crossover between reflexes of PMC *s and *t, but not necessarily in the same etyma. KSR sr is the expected reflex of PMC *t'. The synchronic status of y in KSR is not clear. Some y's may be phonemic, while others seem to be either prothetic or epenthetic. We have not yet arrived at an explanation for these splits.

7.6 Grammatical reconstruction

In the process of splitting up the large cognate sets containing several different forms of the same verbal root (see section 4.6) we accumulated evidence that might prove useful toward the reconstruction of the PMC grammatical system.

Transitives are formed by the addition of a transitivity suffix, usually -i (see example 4.6). These suffixes remain productive in some of the daughter languages. In other languages they are no longer productive, but their former presence has left its trace in the presence of final vowels on transitive verbs — only the suffixal vowel was lost in the historical process of final vowel deletion. So-called 'thematic consonants' have been preserved in a number of forms by the former presence of the transitive suffix — these are marked in example 4.6 by a preceding hyphen. There is evidence that final consonants were deleted, too, in Micronesian languages, if there was no vowel (which may have been deleted subsequently) to protect them.²⁶ The Micronesian evidence for transitivity *-i is reinforced by Pawley's (1973) reconstruction of the same form in Proto-Oceanic.²⁷

All Micronesian languages have causatives. Forms such as KIR ka-, MRS ka-, PNP ka-, MOK ka-, TRK a- (reduplicated as kka-), WOL ga-, and PUA ka- point toward a PMC form *ka-. KSR is peculiar in having ahk- as the productive causative prefix, although MRS does preserve fossilised yak- (< *faka-?) and ya- (< *fa-?). Whether KSR ahk- is *faka- with the regular zero reflex of *f and sporadic loss of *a at the end of the prefix, or a metathesised reflex of *ka-, is not clear. What is unusual about the KSR causative is not only the form of the causative prefix, but also the presence of a transitivity suffix -ye on these forms (Lee 1975:187-189). Lee (pp.178-183) states that -i is used to change certain types of noun and adjectives into transitive verbs. -ye occurs only when the causative prefix is present. Its historical antecedents are not clear.

The other transitive suffix in KSR, -khin (Lee 1975:183-186), presumably reflects a reconstructed 'remote transitive' suffix *aki(ni) (Pawley 1973, Pawley and Reid 1979). This suffix is also found as KIR -akina, MRS ke'n and -Vk, PNP ki or kin (the latter when the following word begins with a vowel), PNG kin, PUL -(y)akin or -(y)ekin, in Carolinian as -ghili, -ghini, or -gini (depending on the dialect), WOL (y)agilli, ULI yixili (-xili as a suffix on some verbs), and PUA akini. MOK -ki reflects proto *-aki.²⁸

Our data also permit the reconstruction of an ancestral passive suffix PMC *-aki which is reflected as, e.g., KIR -aki, MRS -ak/-e'k, KSR -yuhk, MOK -ek, CRL -ágh, and WOL -ag/-eg. Harrison (1982:202) also derives this from

POC *aki(ni). He suggests that this agentless passive suffix may be a Micronesian innovation.

Reduplication also seems to have been present in PMC. Complete reduplication of former CVCV forms is found throughout the Micronesian family. In some languages, e.g., Trukic and Ponapeic, the reduplicated forms reflect medially the final vowel of the reconstructed PMC form, suggesting either that the forms are frozen or that a final vowel is still present in synchronic underlying forms. In Kosraean, on the other hand, complete reduplications have the form CVCCVC, which suggests that such derivations must have occurred after the historical final vowels were lost from the simple forms, or that reduplication was performed on an abstract root that did not contain them.

Micronesian languages appear generally to have formed intransitive verbs by a process of initial CV- reduplication which is reflected synchronically by initial geminate consonants in the Trukic languages and Marshallese, following loss of the vowel between like consonants (Goodenough 1963). The same is true for the more sonorous consonants (e.g., m and ng) in Ponapean, but initial geminate n, l, and r have been reduced to single consonants, and the first members of geminate obstruents have undergone nasal substitution, yielding homorganic nasal-oral consonant clusters, which are preceded by a prothetic high vowel that agrees in rounding with the following segment(s). What appear to be frozen traces of a similar process of initial CV(C)- reduplication – but without syllable reduction – can be found in KSR. Forms still showing this initial CV- reduplication unaltered also can be found in KIR.

The presence of at least the residue of this process in most or all of the daughter languages points to its existence in the ancestral language. Both complete (CVCV) and initial CV(C)- reduplication appear to have existed in PMC. Kiribati uses the latter, and sometimes the former, to form distributive verbs often glossed "abounding in N" or "frequentative of V" in Bingham (1908). The most regular means of forming such verbs in Marshallese is by a combination of the two types of reduplication. So, for example, kkarjinjin *reek of kerosene* from karjin *kerosene*. Kosraean, Mokilese, and Woleaian (and probably many of the other MC languages) also allow final syllable reduplication.²⁹

The existence of a number of reconstructions beginning in *ma- points toward a stative prefix *ma- which may or may not have been productive in PMC, but appears not to be productive in any of the present-day daughter languages.

A word of caution does need to be said on the subject of reconstructing grammatical morphemes in isolation from the rest of the grammatical system. Harrison (1982:181) points out that "slavish devotion" to the principle of reconstructing a unique innovation in the case of shared development in a number of related languages "can easily lead to gross errors in grammatical reconstruction". This he attributes to the practice of applying methods designed for phonological and lexical reconstruction to the reconstruction of grammatical systems. In particular, he claims that POC *aki(ni), which he suggests was a lexical verb appearing in serial construction with a preceding verb, underwent the change to a suffix l) at different times in different branches of the Oceanic language family, and 2) at different times in different functions – these changes being interrelated with other changes in the grammatical system of each daughter language. This he feels is the only way in which one can account for the multiplicity of functions associated with present-day reflexes of this proto-form.³⁰

8. THE GENETIC STATUS OF PROTO-MICRONESIAN

8.1 Introductory

It is generally recognised that the Polynesian languages constitute a very clearly defined family. There may be some disagreement over the exact number of languages (as opposed to dialects) in the family, but there is no language whose membership is at all open to dispute, and the internal subgrouping is relatively well understood. The existence of a close external relationship between Polynesian and Fijian is also quite evident.

The situation in Micronesian is not nearly so clear. Bender (1971) had to include a "questionably nuclear" category for Yapese and Nauruan. We are still uncertain about the genetic affiliation of these two languages. Internal subgrouping is very shallow. External relationships are also unclear.

8.2 Status of Micronesian as an exclusive (?) subgroup

Phonological and grammatical innovations exist which distinguish Proto-Micronesian from Proto-Oceanic, but none have yet been found that are shared by all and only the nuclear Micronesian languages. Nor have any uniquely shared lexical innovations been established to date. No exhaustive search has been made for morphological or semantic innovations.

The limited amount of work that has been done on comparative and reconstructed Micronesian grammar by Harrison (1973, 1978), Sohn (1973), and Sugita (1973) suggests a system that bears considerable resemblance to that reconstructed by Pawley (1973) for Proto-Oceanic. The so-called numeral classifiers do appear to be a Micronesian development, although possibly not uniquely shared³¹ and not equally distributed among the MC languages. In particular, Marshallese has only vestiges, and Kosraean would appear to have missed out altogether in this respect. KSR does have two sets of numbers, but nothing like the elaborate classificatory systems of some of the other languages. Harrison (1976:95-97) describes four classifiers in Mokilese, while Rehg (1981: 124-137) cites some thirty classifiers in Ponapean. Kiribati appears to have the largest number of known classifiers; Trussel (1979:appendix) lists 66 numeral classifiers.

8.3 Internal relationships

Internal relationships within Micronesian, too, are less clear than one might like. The Trukic languages/dialects are obviously closely related. E. Quackenbush (1968) describes them in terms of a dialect chain or continuum. Jackson (1984) gives a list of phonological innovations shared by the Trukic languages which he claims are not shared — as a combined set, although individual innovations may be shared — by any other language or language group:

- (1) Loss of POC *p before round vowels
- (2) Loss of POC *ŋk in all environments
- (3) Loss of POC *q in all environments
- (4) Merger of POC *n with *ŋ in the environment /a ___ i
- (5) Merger of POC *n and *ñ elsewhere
- (6) Merger of POC *s, *ns, and *j

- (7) Separate reflex of (POC?) *nj
- (8) Merger of POC *nt and *nd
- (9) A unique pattern of loss of *R and/or merger with *d
- (10) Loss of POC *y.

Rehg (1981:7-12) indicates that Ponapean and the languages of the nearby atolls — Mokil, Ngatik, and Pingelap — are all mutually intelligible, although the exact degree of intelligibility may vary. Not enough is known about Ngatikese to enable anyone to say much more than that it appears very similar to Ponapean. The 100-word list yields the following percentages of shared cognates among the other Ponapeic languages:

1. PNP - MOK	73%
2. PNP - PNG	79%
3. PNG - MOK	83%

all of which are well above the cognacy rates with the other Micronesian languages. Rehg (pp.9-11) cites other evidence for a closer relationship between PNG and MOK, suggesting the possibility that these may be considered dialects of each other, with Ponapean regarded as a distinct language.

At a slightly higher level, Dyen (1965) proposed a Trukic-Ponapeic subgroup on the basis of lexicostatistical evidence. This evidence ought to be reexamined, however, in the light of our present knowledge of these and other Micronesian languages.

Harrison (1982) cites two apparent innovations that might be used as grammatical evidence to argue for a subgroup consisting of Kosraean and the Ponapeic languages: 1. the use of reflexes of POC *aki(ni) in an instrument-flagging function, e.g.,

- (8.1) KSR Nga owokihn sop ah.
I washed with soap.
- MOK Ih pihnki parnijjo.
He's painting with the varnish.

(pp.204-206), and 2. the use of reflexes of POC *aki(ni) to derive denominal and deadjectival/stative transitives, e.g.,

- (8.2) KSR Eltahl sengseikihn kom.
They consider you a teacher.
- MOK Ngoah jamanki woallo.
I regard that man as a father.

(pp.212-213). Both of these functions of reflexes of POC *aki(ni) appear to be restricted to Ponapeic and Kosraean. However, Harrison (personal communication) does not consider this adequate basis for hypothesising the existence of a Ponapeic-Kosraean subgroup.

Until stronger evidence can be offered for higher-level subgroups, then, we continue to follow Bender's (1971) five-way subgrouping: Kiribati, Marshallese, Kosraean, Ponapeic, and Trukic.

As Marck (1977) pointed out, this apparent shallowness of internal subgrouping makes it more difficult to decide when one is justified in reconstructing a particular form as belonging to Proto-Micronesian rather than to some lower-level proto-language. It would be unreasonable to require that cognate forms occur in all five putative major branches before we attribute the reconstruction to PMC. Marck compromised by using a single star for forms

which were reflected in either Ponapeic or Trukic and at least two of the other three branches, since the probability is relatively high that these forms occurred in PMC, while other reconstructions were marked by double stars to indicate that the distribution of reflexes was defective. Marck reconstructed such forms when he deemed them important for examining the histories of particular languages.

In some cases these double star forms have cognates outside Micronesia. Marck did not distinguish these from other double star forms since he was primarily concerned with the evidence from within Micronesia. Judging from the practices of other comparativists, however, we may consider ourselves justified in reconstructing a form as PMC if we find MC cognates for a POC etymon, even if the form is not found in all major branches of Micronesian, on the grounds that this may be a retention from an earlier stage. The only other plausible explanation for such a distribution is borrowing from outside the MC family. Such borrowings are known to have occurred in languages such as KIR (< PN) and CRL/CRN (< Chamorro). In both instances the borrowings are easily identified.³² In the other MC languages, however, there is no clear evidence of borrowing from non-MC Austronesian languages. It has been suggested that such borrowing may have occurred in KSR, but the suggestion remains unproven.

In general, we have followed Marck's lead and reconstructed Proto-Micronesian forms for cognate sets that included forms from at least three — and preferably four — branches, of which one had to be Trukic or Ponapeic. A number of PTK forms have been reconstructed by Jackson on the basis of exclusively Trukic cognate sets. In some cases, however, a primarily Trukic cognate set may include one non-Trukic witness. We have tentatively reconstructed PMC forms for these sets in the expectation that further search may turn up other non-Trukic reflexes.³³

8.4 External relationships

It is, of course, possible to reconstruct a plausible ancestral language for a set of languages which have not been shown to constitute a valid subgroup. Blust (1984) asserts that Levy (1979) has, in fact, done so for the languages of the South-East Solomons, conglomerating together two groups of languages which Blust believes do not form an immediate subgroup. Instead, Blust presents a case for an immediate subgrouping connection between the Cristobal-Malaitan languages of the South-East Solomons and the languages of nuclear Micronesia.

Both Pawley (1972) and Levy (1979) have presented evidence for the existence of a Cristobal-Malaitan language group. Blust is willing to assume the existence of a nuclear Micronesian subgroup: nonetheless, he is careful to state that his hypothesis would not be seriously damaged if nuclear Micronesian were shown not to be an exclusive subgroup. Blust's difficulty is that he can find no phonological or grammatical innovations that are uniquely shared between Cristobal-Malaitan and nuclear Micronesian languages. Loss of POC *q and of original final consonants, cited by Blust as characteristic of both Malaitan and Micronesian languages, are shared with a number of other Oceanic languages.

Blust's evidence for a Malaitan-Micronesian subgroup falls into three categories: 1) lexical, 2) morphological, 3) semantic. All of these are subject to the danger that shared innovations may be difficult to distinguish from shared retentions. Blust has made what appears to be an exhaustive check of the available data from other Oceanic languages in order to minimise the

'avoidable error' of failing to discover existing external cognates. There is, however, no way to avoid the error that may occur if a retention should happen to be restricted to just the set of languages that constitute one's putative subgroup, or in the case of morphological and semantic innovations, if drift has occurred in the same direction only among those languages. Blust suggests, however, that while coincidences may occur on occasion, a large number of such coincidences is not likely unless these languages do, in fact, form a subgroup. Thus, he takes the number of common features — even though these cannot be shown definitively to be uniquely shared innovations — to be a kind of sub-grouping evidence, nonetheless. The numbers are relatively small, however; one wishes that more data were available from a larger number of languages.

Aside from Blust's Malaitan-Micronesian connection, attempts have also been made to link the nuclear Micronesian languages with the languages of the Admiralty Islands (Smythe 1970) and with the North Hebridean - Central Pacific languages (Pawley 1972). Both of these are rather thoroughly demolished by Blust.³⁴

8.5 Speculations on Micronesian prehistory

The archaeological evidence for Micronesia is scanty, but Cordy (1982) cites a number of dates in the 400 B.C. (Truk lagoon) to A.D. 400 range for the Marshalls, Ponape, Truk lagoon, Ulithi, and Yap. No early dates are available for Kosrae. Cordy feels that sites earlier than the ones reported on should exist in Ponape, Kosrae, and the Marshalls, but just have not been discovered yet.

The available dates suggest relatively rapid dispersal and settlement, but do not give any particular indication as to direction of dispersal, whereas the linguistic evidence points to a dispersal from the east, the area of greatest linguistic diversity being in eastern Micronesia — the Marshalls, Kiribati, Nauru, Kosrae, Ponape, and Truk lagoon.

The earliest archaeological dates available for geographic Micronesia are from the Marianas. These are considerably earlier than the dates for the eastern islands. The material — as well as linguistic — evidence points toward settlement of the Marianas by a different group of people coming from the other direction, possibly from the Philippines.

These people may have explored the islands to the east as well, but if they did settle them, it was considerably later than their settlement of the Marianas — and the settlements may not have persisted, or else may have been absorbed by later-arriving nuclear Micronesian speakers. In all probability, these early western Micronesians found the atolls inhospitable and never did get as far as the high islands at the eastern end of the Carolines.

The earliest dates cited by Cordy for the western Caroline atolls of Ulithi and Ngulu are in the same range as the eastern island sites, as are early dates from Palau (and Yap?). The Palauan language, at least, appears to be more closely affiliated with western Austronesian languages than with nuclear Micronesian. The linguistic affiliations of Yapese have not been established, although Bender (1971) is willing to admit Yapese as Oceanic and possibly even nuclear Micronesian.

It is possible that detailed examination of the sound correspondences in our present data may lead to the discovery of shared innovations that would

allow us to determine the relationships among the various branches of nuclear Micronesian. If archaeologists could get funding for excavations rather than just site surveys and salvage work, they might find more evidence bearing on settlement and origins. At present, however, the only prehistoric sequence of any sort available for anywhere in Micronesia is for the Marianas.

Of course, it may be that the Proto-Micronesians — even assuming that there was only one group of original settlers who all came from the same 'homeland' — spread so rapidly across nuclear Micronesia that there was virtually no period of common Micronesian development, and thus no uniquely shared linguistic innovations, either within nuclear Micronesian or between major branches of the putative subgroup. Unlike Polynesia, which consists largely of geographically distant island groups, the islands of nuclear Micronesia lie relatively close together. Certainly people who, in Blust's view, were capable of making the long voyage north from the South-East Solomons would have had no difficulty sailing from one Micronesian island to another.

In that case, the ancestral language of the nuclear Micronesians may not have been Proto-Micronesian, but Proto- _____ -Micronesian. Blust has suggested that the blank should be filled by 'Malaitan'. It behooves us, as proponents of the nuclear Micronesian hypothesis, to examine Blust's proposal carefully — and also to reexamine the other subgrouping hypotheses that he alludes to (Smythe 1970, Pawley 1972) — to make comparisons with the languages that have been put forward as immediate relatives of the nuclear Micronesian languages, and perhaps in that way discover what, if anything, sets nuclear Micronesian apart from the rest.

NOTES

1. Neither this report nor the project as a whole would have been possible without the effort of a number of people. The core group consisted of Byron W. Bender, Robert W. Hsu, Frederick H. Jackson, Jeffrey C. Marck, Kenneth L. Rehg, Ho-min Sohn, Stephen Trussel, and Judith W. Wang. Some of these people are no longer at the University of Hawaii; all have made significant contributions to the project. We also gratefully acknowledge the contributions of visiting colleagues from other institutions: Paul Geraghty, Ward H. Goodenough, and Sheldon P. Harrison. A number of graduate students have lent temporary assistance to the project, including Martin Combs, Layla Ebrahim, Elaine Good, Gregg Kinkley, and Meryl Siegal. Sue Archibeque did much of the initial data entry.
2. Or twelve, depending on the status assigned to Saipan Carolinian.
3. See Goodenough and Sugita (1980) and Jackson et al. (to appear) for details.
4. Sohn (1975) does not say it in so many words, but it appears that the Woleaian alphabet is essentially phonemic, with the exception of certain geminate consonants and possibly also the long vowels represented by eo and oa, and the ambiguity of the digraph iu with respect to vowel length.

Historically, eo and oa come from sequences of vowels, but they are simply long vowels in the modern language (p.18).

5. See Elbert (1974) for a description of the phonemic inventory of Puluwatese. Nowhere in either work does Elbert discuss spelling conventions.
6. See Oda (1977:9) for the phonemic inventory of PUA; Oda states in the prefatory material to the dictionary/appendix that SNS forms are generally identical to PUA except that SNS distinguishes two fricative phonemes /f/ and /d/, which have merged as /d/ in PUA.
7. H. Quackenbush (1970) describes the phonemic inventories of four Trukic dialects: Moen (Truk lagoon), Pullap, Satawal, and Sonsorol. According to Quackenbush, Moen, Pullap, and Satawal all have the same nine-vowel inventory which is also found in Saipan Carolinian. Seven vowels are given for Sonsorol — Quackenbush uses a different set of symbols and arranges them slightly differently, but basically agrees with Oda. Quackenbush's reconstructed Proto-Trukic has six vowels, the same six found in Marck's reconstructed Proto-Micronesian. Woleaian has the same six vowels in its short vowel set, but has two additional vowels in the long vowel set (see note 4).
 All of the Trukic languages/dialects have similar consonantal inventories, differing from such eastern languages as Marshallese and Kosraean in not having three different sets of consonant types — plain, velarised, and labialised. In Trukic, the only distinctively velarised consonants are w and the labiovelars pw and mw.
8. Or, from a surface point of view, four. See Bender (1968) for details.
9. The symbols used in this chart are the standard orthographic representations.
10. See Hsu and Peters (1984) for a description of the development of dictionary-processing by computer at the University of Hawaii.
11. Additional blank spaces have been inserted in this and other examples for legibility. These are not present in the data as entered into the computer, but would be inserted by the DISPLIGN program. See Section 5.1.
12. The question of *s and *S (S' in the computer orthography) in Proto-Micronesian is actually more complicated than Marck realised. See section 7.5.
13. Eventually to be replaced by KIR.
14. Actually, the dots do get picked up by the concordance program, but this does not create any additional complications in the correspondences because the program is designed to generate dots as placeholders when there is no cognate form in a given language.
15. See also POC *RapiRapi in example 3.2.
16. Both this inconsistency and the inconsistent use of X for both cognate and non-cognate forms must be attributed to the inefficiency inherent in such an informally-organised project. Many of the conventions and practices that we describe here are not the outcome of careful deliberation, but are the result of spur-of-the-moment decisions made by the person entering the data into the computer. These and a number of other aspects of the file will have to be revised before any publication of the results of this project.

17. See section 7.5 for one example of this phenomenon.
18. Actually, DISPLIGN inserts more blanks than are shown. This is due to the relative narrowness of the paper used here. DISPLIGN printouts normally appear on wide (15" x 11") paper.
19. This ordering can be changed to some other order specified by the user.
20. The vowel correspondences pose an entirely different – and quite difficult – problem, which will not be discussed here.
21. The periods ('dots') in the *t₃ and *t₄ correspondences represent gaps where cognates are lacking. The quantities given for these correspondences obviously do not represent full correspondences but occurrences of the given partial correspondence.
22. These are reproduced in slightly altered form in order to fit them onto the page. ALIGN printouts normally appear on wide (15" x 11") paper which allows space for line numbers, longer glosses, and more white space for greater legibility. See the sample concordance page given as Appendix C.
23. Although the three full *s₂ correspondences are not augmented by including other correspondences with gaps, one of the three is noteworthy in being a doublet of a regular *s correspondence, because of MRS alternants: PMC *s₂ama *outrigger* – TK *tama *outrigger float*, PNP dɛmɛ *outrigger* (3ps), MRS yam *sail with outrigger out of water*, KIR rama *outrigger* (cf. MRS (rey-)tam *the outrigger side of a canoe* in the regular *s correspondence).
24. There were no full correspondences with cognates in all four languages. These correspondences are represented only by overlapping sets of correspondences with gaps.
25. Numbers of occurrences in this table may disagree with those in 7.4. – N here is a total including non-conflicting correspondences with gaps.
26. See Lee and Wang (1984) for examples of such developments in Kosraean.
27. See Harrison (1978) for a fuller discussion of transitivity in Micronesian languages.
28. See Harrison (1982) for a different interpretation of POC *aki(ni).
29. See Harrison (1973) for further discussion of reduplication in Micronesian.
30. See examples in part 1.2, pp.179-180.
31. This has not yet been established.
32. As are such intra-Micronesian borrowings as MOK jimwoa *chicken* < KIR te moa, and various Marshallese terms associated with coconut toddy (Bender 1981).
33. Because of the dispersed nature of the Micronesianist group in the past few years, the most recent additions to the file have not been as thoroughly checked – either for the existence of cognates or for accuracy of data entry – as were earlier data. It is probable that many of the gaps in the data will be filled when all languages have been carefully checked.
34. In fairness to Pawley, it ought to be said that he has since revised his Eastern Oceanic subgroup a couple of times (1977, 1979), so that it no longer includes either the South-East Solomonic or the nuclear Micronesian languages.

FAKA	AEI	46	AD	PA	RAPI	EVENING			
FAKA	AFI	43	AN	RA	ABI	EVENING			
A	FARA	25	AN	BAR	A	SHOULDER			
FA	O	66	AN	BA	QER	U (O) NEW			
FA	O	60	CRL	(-)	FE'	NEW			
FAKA	AFI	36	CRL	(LEE-)	F A'	FF - EVENING			
FAKA	AFI	37	CRL	(LEE-)	F A'	A' F - IN THE EVENING			
A	FARA	18	CRL	A (Y)	U' FARA (-)	SHOULDER (& AIFARA)			
A	FARA	20	CRL	A I	FARA (-L)	HIS SHOULDER (& AFAR)			
KINI		96	CRL	GH	I L I (-)	PICK UP, GATHER			
KINI	> KINI -T (-I)	122	CWL	GH	I L I (-I)	VT. TO PICK UP, HARVEST			
KAKANGI		86	CRL	K	K A' NG	SHARP			
MAMATA		143	CRL	M	M A S A (-TA')	WAKE UP			
A	FARA	19	CRLX	A I	FAR	SHOULDER (& AFARAL)			
MAMATA		142	CRLX	M	M A S	AWAKE			
FAKA	AEI	67	ED	RA	AVI	EVENING			
FA	O	70	ED	V	A QAR	U NEW			
MAMATA		148	FIJ	M	MATA	ID			
FAKA	AEI	44	FIJ	(MATA-)	A K	A V I EVENING			
FA	O	72	FIJ	V	D	U NEW			
FAKA	AFI	49	FIJ	Y	A K	A V I EVENING			
KAKANGI		75	GIL	K	A K A NG	SHARP			
KINI	> KINI -T (-I)	113	GIL	K	I N I -K (-A)	VT. PINCH			
KINI	> KINIKINI	104	GIL	K	I N I K I N	N. PINCH			
FA	O	52	GILL	B	O	O U NEW (<TUV FOOU)			
MAMATA		132	GILX			(MAMATA = INTUITION, SAGACIOUS)			
KAKANGI		78	KSR	(LAH-)	K UH NG	SHARP, SMART, INTELLIGENT			
FAKA	AEI	29	KSR	E	K UH	EVENING			
KINI		94	KSR	K	I H N	VI. PINCH, PICK			
KINI	> KINIKINI	105	KSR	K	I H N	K I H N VI. PINCH, PICK			
KINI	> KINIKINI -T (-I)	115	KSR	K	I H N I -S	PINCH, PICK			
A	FARA	2	MC	A	FARA	*SHOULDER (<*-BOD>)			
FA	O	51	MC	F	A	O	U *NEW (QB 122)		
FAKA	AEI	28	MC	F	A	K	A	F I *EVENING (<*-MET <*-TIM>)	
KAKANGI		74	MC	K	A	K	A	N G I *SHARP (<*-CUT>)	
KINI		93	MC	K	I	N I	VI. *PINCH, *PICK (<*-CUT>)		
KINI	> KINI -T (-I)	112	MC	K	I	N I	-T (-I) VT. *PINCH, *PICK (<*-CUT>)		
KINI	> KINIKINI	103	MC	K	I	N I	K (N I) VI. *PINCH, *PICK (<*-CUT>)		
MAMATA		131	MC	M	A	M	A	T A *AWAKE (<*-PHY>)	
KAKANGI		79	MOK	(I)	NG	K	D	A	NG SHARP
A	FARA	6	MOK	A	P	R	O	A	SHOULDER (3PS)
KINI	> KINI -T (-I)	116	MOK	K	I	N I	VT. PINCH WITH FINGERNAILS		
KINI	> KINIKINI	106	MOK	K	I	N I	K I N	PINCH WITH FINGERNAILS	
A	FARA	4	MRS	H	A	Y	E	R	A (-), SHOULDER (COMBINING FORM FOR PER. POSS. SUFFIXES) (& HAYERAT)
KAKANGI		76	MRS	K	K	A	G	SHARP (& KKAGKE*G)	
KINI	> KINI -T (-I)	114	MRS	K	I	N	-J (-IY)	PINCH WITH FINGERNAILS	
MAMATA		133	MRS	M	M	E	J	KEEP AWAKE, STAY UP	
A	FARA	3	MRSX	H	A	Y	E	R	A (V) SHOULDER
KAKANGI		77	MRSX	K	K	A	G	(-KE*G) SHARP (POSTPOSITIONAL FORM) (& KKAGL	
MAMATA		136	MRT	(A-)	M	M	A	S	A' (A'TA') VT. WAKE HIM
FAKA	AFI	31	MRT	(LE-)	F	A'	A' F	EVENING	
FAKA	AEI	32	MRT	(LEE-)	F	A'	A' F	IN THE EVENING	
A	FARA	11	MRT	A (W)U'	F	A	R	(-N) SHOULDER (3PS)	
FA	O	55	MRT	F	E'	NEW			
KAKANGI		82	MRT	K	K	A	NG	SHARP	
KINI	> KINI -T (-I)	119	MRT	K	I	N I	(-Y) BREAK PIECES OFF, CUT		
KINI	> KINIKINI	109	MRT	K	I	N I	K I N	ACTION OF CUTTING OR BREAKING OFF PIECES	
FA	O	54	MRTX	(SE'E')	F	E'	NEW		
A	FARA	12	MRTX	A (W)U'	F	A	R	SHOULDER	
MAMATA		135	MRTX	M	M	A	S	VI. GUARD	

Appendix A: BANDSORT sample

FINDERLIST FOR MIC.SAMPLE

00000420

COMPILED 01/13/83

1	AWAKE	
2		M_NATMA <TK>:: *
3		MAMATA <MC>:: *
4	BREAK	
5		KINIKINI <TK>:: PINCH OFF, BREAK OFF (AS PIECES).
6	EVENING	
7		FAKA_AFI <MC>:: *
8		FAKA_AFI <TK>:: *
9	HARVEST	
10		KINI-T*(-I) <TK>:: PLUCK, HARVEST.
11	NEW	
12		FA_O_U <MC>:: NEW (OB 122).
13		FA_O_U <TK>:: NEW (OB 122).
14	PICK	
15		KINI <MC>:: VI, PINCH, PICK.
16		KINI <TK>:: PICK UP, PLUCK (E.G., FRUIT).
17		KINI-T*(-I) <MC>:: VI, PINCH, PICK.
18		KINIKINI <MC>:: VI, PINCH, PICK.
19	PINCH	
20		KINI <MC>:: VI, PINCH, PICK.
21		KINI-T*(-I) <MC>:: VI, PINCH, PICK.
22		KINIKINI <MC>:: VI, PINCH, PICK.
23		KINIKINI <TK>:: PINCH OFF, BREAK OFF (AS PIECES).
24	PLUCK	
25		KINI <TK>:: PICK UP, PLUCK (E.G., FRUIT).
26		KINI-T*(-I) <TK>:: PLUCK, HARVEST.
27	SHARP	
28		K_KANGI <TK>:: *
29		KAKANGI <MC>:: *
30	SHOULDER	
31		A_FANA <MC>:: *
32		-A_FARA <TK>:: SHOULDER (& AUFARA).
33		-AUFANA <TK>:: SHOULDER (& AFARA).
34	WAKE	
35		MAMATA <MC>:: AWAKE.

END OF LISTING.
 35 LINES, 11 MAIN ENTRIES, 0 SUBENTRIES.
 24 RECORDS IN SORTED FILE, 11 DIFFERENT KEYWORDS.
 LENGTH OF LONGEST PARAGRAPH SET: 54

Appendix C: concordance data sample

MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	(I) - . - -	PAGE 3
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		(I) -	
MIC.SAMPLE.880 *SHARP <*-CUT>	TK	K - K A N G I	-	
MIC.SAMPLE.870 SHARP, OF AN EDGE OR POINT	PNP	* - K A N G -	-	
MIC.SAMPLE.830 SHARP (& KKAGKE'G)	MRS	K - K A G -	-	
MIC.SAMPLE.820 SHARP	GIL	K A K A N G -	-	
MIC.SAMPLE.850 SHARP, SMART, INTELLIGENT	KSR	(LAHL-) * - K J H N G -	-	
MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	(I) I . I -	
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		(I I) I I I H	
MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	-	
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		2	
MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	-	
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		(I I) - I I	
MIC.SAMPLE.1250 *PLUCK, *HARVEST <*-CUT>	TK	K I N I -T- (-I)	-	
MIC.SAMPLE.1240 VT. PINCH	PNP	K I N I -J (-IY)	-	
MIC.SAMPLE.1210 PINCH WITH FINGERNAILS	MRS	K I N I -K (-A)	-	
MIC.SAMPLE.1200 VT. PINCH	GIL	K I H N I -S	-	
MIC.SAMPLE.1220 PINCH, PICK	KSR		(I I) I I I H	
MIC.SAMPLE.1250 *PLUCK, *HARVEST <*-CUT>	TK	K I N I -T- (-I)	-	
MIC.SAMPLE.1240 VT. PINCH	PNP	K I N I -J (-I)	-	
MIC.SAMPLE.1210 PINCH WITH FINGERNAILS	MRS	K I N I -K (-A)	-	
MIC.SAMPLE.1200 VT. PINCH	GIL	K I H N I -S	-	
MIC.SAMPLE.1220 PINCH, PICK	KSR		K . . . K	
MIC.SAMPLE.300 *EVENING <*-MET *-TIM>	TK	F A K A A P I	-	
	PNP		-	
	MRS		-	
	GIL		-	
	KSR		E K UH	
MIC.SAMPLE.290 EVENING				
MIC.SAMPLE.1020 *PICK UP, *PLUCK (E.G., FRUIT) <*-CUT>	TK	K I N I	-	2
	PNP			
	MRS			
	GIL			
	KSR			
MIC.SAMPLE.1010 VI. PINCH, PICK				
MIC.SAMPLE.880 *SHARP <*-CUT>	TK	K I H N	(K .) K K .	
MIC.SAMPLE.870 SHARP, OF AN EDGE OR POINT	PNP	K - K A N G I	-	
MIC.SAMPLE.830 SHARP (& KKAGKE'G)	MRS	K - K A G -	-	
MIC.SAMPLE.820 SHARP	GIL	K A K A N G -	-	
MIC.SAMPLE.850 SHARP, SMART, INTELLIGENT	KSR	(LAHL-) * - K J H N G -	-	
MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	(K) K . K K	
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		2	
MIC.SAMPLE.1150 *PINCH OFF, *BREAK OFF (AS PIECES) <*-CUT>	TK	K I N I K I N I	-	
MIC.SAMPLE.1140 PINCH	PNP	K I N I K I N I	-	
MIC.SAMPLE.1110 N. PINCH	MRS	K I N I K I N I	-	
MIC.SAMPLE.1120 VI. PINCH, PICK	GIL	K I H N - K I H N	-	
	KSR		(K K) K K K	
MIC.SAMPLE.880 *SHARP <*-CUT>	TK	K - K A N G I	-	
MIC.SAMPLE.870 SHARP, OF AN EDGE OR POINT	PNP	* - K A N G -	-	
MIC.SAMPLE.830 SHARP (& KKAGKE'G)	MRS	K - K A G -	-	
MIC.SAMPLE.820 SHARP	GIL	K A K A N G -	-	
MIC.SAMPLE.850 SHARP, SMART, INTELLIGENT	KSR	(LAHL-) * - K J H N G -	-	
			(K K K K K)	

BIBLIOGRAPHY

- ABO, Takaji, Byron W. BENDER, Alfred CAPELLE, and Tony DEBRUM
 1976 *Marshallese - English dictionary*. Honolulu: University Press of Hawaii.
- BENDER, Byron W.
 1968 Marshallese phonology. *Oceanic Linguistics* 7/1:16-35.
 1971 Micronesian languages. In Thomas A. Sebeok, ed. *Current trends in linguistics*, vol.8, Part 1: *Linguistics in Oceania*, 426-465. The Hague: Mouton.
 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: The Linguistic Society of New Zealand.
- BENDER Byron W., ed.
 1984 *Studies in Micronesian linguistics*. PL, C-80.
- BINGHAM, Hiram
 1908 *A Gilbertese-English dictionary*. Boston: American Board of Commissioners for Foreign Missions.
- BLUST, Robert
 1976 A third palatal reflex in Polynesian languages. *Journal of the Polynesian Society* 85:339-358.
 1978 *The Proto-Oceanic palatals*. Wellington: The Polynesian Society.
 1984 Malaita - Micronesian: An eastern Oceanic subgroup? *Journal of the Polynesian Society* 93/2:99-140.
- CORDY, Ross
 1982 A summary of archaeological work in Micronesia since 1977. *Bulletin of the Indo-Pacific Prehistory Association* 3:118-128.
- DYEN, Isidore
 1949 On the history of the Trukese vowels. *Language* 25:420-436.
 1965 *A lexicostatistical classification of the Austronesian languages*. IJAL Memoir No. 19. Indiana University Publications in Anthropology and Linguistics.
- ELBERT, Samuel H.
 1972 *Puluwat dictionary*. PL, C-24.
 1974 *Puluwat grammar*. PL, B-29.
- GOODENOUGH, Ward H.
 1961 Migrations implied by relationships of New Britain dialects to Central Pacific languages. *Journal of the Polynesian Society* 70:112-136.

GOODENOUGH, Ward H.

- 1963 The long or double consonants of Trukese. In the *Proceedings of the Ninth Pacific Science Congress, 1957*, vol.3:77-86.

GOODENOUGH, Ward H. and Hiroshi SUGITA

- 1980 *Trukese-English dictionary*. Philadelphia: American Philosophical Society Memoirs 141.

GRACE, George W., Janet BLACK, Peter C. LINCOLN, and James THARP

- n.d. Proto-Oceanic lexical reconstructions. Computer printout (1978). Honolulu: Department of Linguistics, University of Hawaii.

HARRISON, Sheldon P.

- 1973 Reduplication in Micronesian languages. *Oceanic Linguistics* 12/1-2: 407-454.
- 1976 *Mokilese reference grammar*. Honolulu: University Press of Hawaii.
- 1978 Transitive marking in Micronesian languages. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 1067-1127. PL, C-61.
- 1982 Proto-Oceanic *aki(ni) and the Proto-Oceanic periphrastic causative. In Amran Halim, Lois Carrington, and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.1: *Currents in Oceanic*, 178-230. PL, C-74.

HARRISON, Sheldon P., and Salich ALBERT

- 1977 *Mokilese-English dictionary*. Honolulu: University Press of Hawaii.

HSU, Robert W. and Ann M. PETERS

- 1984 Computers and Micronesian dictionaries: a chronicle of systems-design fieldwork among the lexicographers. In Bender, ed. 1984:1-36.

JACKSON, Frederick H.

- 1984 Reflexes of Proto-Oceanic in the Trukic languages of Micronesia. In Bender, ed. 1984:259-280.

JACKSON, Frederick et al.

- f/c *Saipan Carolinian dictionary*.

JENSEN, John Thayer

- 1977 *Yapese-English dictionary*. Honolulu: University Press of Hawaii.

LEE, Kee-dong

- 1975 *Kusaiean reference grammar*. Honolulu: University Press of Hawaii.
- 1976 *Kusaiean-English dictionary*. Honolulu: University Press of Hawaii.

LEE, Kee-dong and Judith W. WANG

- 1984 Kosraean reflexes of Proto-Oceanic phonemes. In Bender, ed. 1984: 403-442.

LEVY, Richard

- 1979 Languages of the Southeast Solomon Islands and the reconstruction of Proto-Eastern Oceanic. In Naylor, ed. 1979:213-226.

LINCOLN, Peter C.

- 1981 Micronesia. In Stephen A. Wurm and Shirô Hattori, eds *Language atlas of the Pacific area, Part I: New Guinea area, Oceania, Australia*. Canberra: The Australian Academy of the Humanities, in collaboration with the Japan Academy; PL, C-66.

LÜTKE, Frederic

- 1835- *Voyage autour du monde, exécuté par ordre de Sa Majesté l'empereur*
1836 *Nicolas Ier sur la corvette Le Seniavine, dans les années 1826, 1828 et 1829*. 3 vols. Paris: Atlas.

MARCK, Jeffrey

- 1977 Preliminary nuclear Micronesian sound correspondences. Paper presented to the Linguistic Society of America Summer Meeting, Honolulu.

MATTHEWS, William K.

- 1950 Characteristics of Micronesian. *Lingua* 2:419-437.

MCCAULEY, David P.

- 1966 *Lessons in Kusaien*. Hoolehua, Molokai: UH Peace Corps Training Center.

NAYLOR, Paz B., ed.

- 1979 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*. Ann Arbor: University of Michigan Center for South and Southeast Asian Studies.

ODA, Sachiko

- 1977 The syntax of Pulo Annian: a nuclear Micronesian language. Ph.D. dissertation, University of Hawaii.

PAWLEY, Andrew

- 1972 On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Readings in Oceanic culture history*, vol.3, Pacific Anthropological Records 13:1-142. Honolulu: Bishop Museum.
- 1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12/1-2: 103-188.
- 1977 On redefining 'Eastern Oceanic'. MS.
- 1979 *New evidence on the position of Rotuman*. University of Auckland Anthropology Department Working Papers. No. 56.
- n.d. A Proto-Oceanic grammar. MS.

PAWLEY, Andrew and Lawrence A. REID

- 1979 The evolution of transitive constructions in Austronesian. In Naylor, ed. 1979:103-130.

QUACKENBUSH, Edward M.

- 1968 From Sonsorol to Truk: a dialect chain. Ph.D. dissertation, University of Michigan.

QUACKENBUSH, Hiroko C.

- 1966 *Lessons in Pullapese*. Honolulu: University of Hawaii Peace Corps Training Center.
- 1970 Studies in the phonology of some Trukic dialects. Ph.D. dissertation, University of Michigan.

REHG, Kenneth L.

- 1981 *Ponapean reference grammar*. Honolulu: University Press of Hawaii.
- 1982 Ponapean possessive paradigms: the smart solution, the dumb solution, and the Ponapean solution. Paper presented to the University of Hawaii Austronesian Circle, Honolulu, April 22, 1982.

REHG, Kenneth L., and Damian G. SOHL

- 1979 *Ponapean-English dictionary*. Honolulu: University Press of Hawaii.

SABATIER, Fr. E., MSC, translated by Sister Oliva, FNDSC.

- 1971 *Gilbertese-English dictionary*. Noumea: South Pacific Commission Publications Bureau.

SMYTHE, W. E. (ed. by A. HEALEY)

- 1970 Melanesian, Micronesian, and Indonesian features in languages of the Admiralty Islands. In S.A. Wurm and D.C. Laycock, eds *Pacific linguistic studies in honour of Arthur Capell*, 1209-1234. *PL*, C-13.

SOHN, Ho-min

- 1973 Relative clause formation in Micronesian languages. *Oceanic Linguistics* 12/1-2:353-392.
- 1975 *Woleaian reference grammar*. Honolulu: University Press of Hawaii.

SOHN, Ho-min and Byron W. BENDER

- 1973 *A Ulithian grammar*. *PL*, C-27.

SOHN Ho-min and Anthony F. TAWERILMANG

- 1976 *Woleaian-English dictionary*. Honolulu: University Press of Hawaii.

SUGITA, Hiroshi

- 1973 Comparison of verb-object relationships in Micronesian languages. *Oceanic Linguistics* 12/1-2:393-406.

TRUSSEL, Stephen

- 1979 *Kiribati (Gilbertese) grammar handbook*. Peace Corps Language Handbook Series. Brattleboro, Vermont: The Experiment in International Living.
- n.d. PMC *t and *k in Gilbertese. MS.

WANG, William S-Y.

- 1979 Language change — a lexical perspective. *Annual Review of Anthropology* 8:353-371.

POSSESSIVE CONSTRUCTIONS IN OCEANIC LANGUAGES AND IN PROTO-OCEANIC

Frantisek Lichtenberk

1. INTRODUCTION

The possessive constructions of Oceanic have been the object of interest of students of Oceanic since the last century.¹ This interest has been both language-specific, descriptive, and comparative and historical. A large majority of the Oceanic languages exhibit multiple types of possessive construction. Studies of the first kind have sought to characterise the possessive systems of individual languages from the point of view of their structures as well as use. Studies of the second kind have looked for cross-linguistic commonalities in the possessive systems. These recurrent features have then been used in reconstructions of the system of possessive constructions of Proto-Oceanic.

The present paper falls in the latter category. Its aim is two-fold. The first is to provide a typology of the possessive constructions found in modern Oceanic languages. The typology is a refinement of earlier works by various investigators, such as Codrington (1885), Ray (1919) and Pawley (1973). The other aim is to offer a reconstruction of the Proto-Oceanic possessive system. The reconstruction presented here is basically an elaboration and extension of the most comprehensive reconstruction of the Proto-Oceanic possessive system to date, that of Pawley 1973. The two aims are not entirely disparate: the typology established for the present-day languages will serve as a framework in the reconstruction.

The data on which the present study is based come from over 200 Oceanic and several non-Oceanic languages.² Needless to say, the quantity and the quality of the information available on the possessive constructions of various languages vary tremendously: from quite detailed and perceptive to fragmentary and obscure.

The paper is structured as follows: After a few general terminological remarks in Section 2, a basic typology of the Oceanic possessive types is given in Sections 3 and 4. In Section 5, the semantics of the various possessive types is discussed. In Section 6, a reconstruction of the Proto-Oceanic possessive system is presented. The paper concludes with a brief discussion of a few other topics that are of relevance to Oceanic possessive constructions.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 93-140. *Pacific Linguistics*, C-88, 1985.

© Frantisek Lichtenberk

2. TERMINOLOGICAL PRELIMINARIES

In the discussion that follows, the terms 'possessive construction', 'possessor' and 'possessed' are used as technical terms. A possessive construction may or may not express true possession, i.e. ownership: my car, meaning *the car I own*, or John's car, meaning *the car John owns*. Possessive constructions, as the term is applied here, are commonly used to express many other types of relationship between two entities, such as part-whole (my head) and kinship (John's father). Similarly, my house may refer to the house I live in even though I do not own it, and John's arrival refers to John's involvement in an event, not to ownership. What all the semantic relationships just exemplified as well as many others have in common is the structures used to express them: a small, well-defined set of noun-phrase constructions.

The term 'possessor' and 'possessed' are semantic-syntactic labels for the forms that in an expression of true possession encode the entity that owns another entity, and the entity owned, respectively. In my car (*the car I own*), my is the possessor, and car is the possessed. Again by extension, the two terms are applied to the two elements found in the corresponding positions in other possessive constructions, regardless of the nature of the semantic relationship expressed. In John's arrival, John is the possessor, and arrival is the possessed.

In English and many other languages, the possessed is the head of the noun phrase and the possessor its attribute. In the subsequent discussion I will use the terms 'possessed' and 'possessor' in preference to 'head' and 'attribute' for two reasons:

1. there is a common type of possessive construction found in Oceanic languages in which the possessor is not a separate element but an inflectional affix, either on the possessed or on a classifier;
2. the two terms will refer not only to linguistic elements but, when convenient, also to their referents. Context should make it clear which of the two meanings is intended.

Finally, the term 'possessive construction', as used here, does not subsume constructions with non-specific/non-referential nouns in attribute position (for a brief discussion of this type of noun phrase see Section 6.5.2).

3. TYPOLOGY OF OCEANIC POSSESSIVE CONSTRUCTIONS BASED ON THE TYPES OF CONSTITUENT AND THE RELATIONS BETWEEN THE CONSTITUENTS

For purposes of presentation, it is convenient to deal with the typology of Oceanic possessive constructions in two parts. The first, which is the subject of this section, considers the types of constituent that make up the constructions and the structural relations that obtain between the constituents. The second, in Section 4, is concerned with the linear order of the constituents.

The main point of these two sections is to discuss and exemplify the various basic, recurrent Oceanic possessive types. Nothing will be said about their distribution in the Oceanic subgroup. This aspect of possessive constructions will be considered when we proceed to reconstructing the Proto-Oceanic system.

3.1 Possessive affixes

Oceanic languages typically exhibit sets of affixes whose function is to encode the possessor, as in (1) below, or to cross-reference the possessor, as in (2):

<p>KWAIO (1) asi -mu <i>younger brother-your (sg)</i> <i>your younger brother</i></p>	<p>MANAM (2) boro ta'e -di <i>pig faeces-their</i> <i>the pigs' excrements</i></p>
---	---

The possessive affixes are of two types: suffixes, as in (1) and (2), and prefixes:

<p>WAYAN (3) m̄ -ulu <i>your (sg)-head</i> <i>your head</i></p>	<p>KALIAI (4) yuŕa ai -mata <i>G. his-eye</i> <i>Gura's eye</i></p>
---	--

The possessive affixes are attached either to the possessed, as in (1)-(4), or to a special possessive classifier (see Section 3.2):

<p>KALIAI (5) e -le yariu <i>his-CLASS shield</i> <i>his shield</i></p>	<p>STANDARD FIJIAN (6) na no -na vale na tūraga art CLASS-his house art chief <i>the chief's house</i></p>
---	---

The term 'possessive affixes' is not entirely satisfactory because in some languages the affixes are used in constructions other than possessive. For example, in Manam the suffixes are also used on demonstratives, one class of adjectives and a few numerals (for similar uses of the suffixes in other New Guinea languages see Bradshaw 1982). In a number of languages (e.g. Houailou), the same set of suffixes is used in possessive constructions to index or cross-reference the possessor and on transitive verbs to index or cross-reference the object. Since I will have nothing to say about the use of these affixes in constructions other than possessive, I will use the terms 'possessive affixes', 'possessive suffixes' and 'possessive prefixes' as convenient labels.

3.2 Direct, indirect and prepositional possessive constructions

Cases of ellipsis apart, every possessive construction contains two basic elements: a possessed and a possessor. Besides these two, a possessive construction may contain a third element, which may be of one of two types: a classifier or a preposition. It is then possible to distinguish three basic types of possessive construction: direct (POSS'D, POSS'R), indirect (POSS'D, POSS'R, CLASS), and prepositional (POSS'D, POSS'R, PREP).³

The basic structure of the prepositional possessive construction is
 POSS'D PREP POSS'R.

<p>BABATANA (7) pade ta re <i>house of you (sg)</i> <i>your house</i></p>	<p>HOUAILOU (8) nevā i mōrua i wi?-a? <i>land of grandfather of man-this</i> <i>this man's grandfather's land</i></p>
--	---

The direct-indirect distinction is the hallmark of the Oceanic subgroup. A direct possessive construction consists of a possessed and a possessor. The possessor may be a possessive affix or a separate word (see Section 3.3 for more detail):

IAAI

- (9) *ie -n*
name-his
his name

ARE

- (11) *manu daguri -na*
bird feather-its
the bird's feather(s)

GILBERTESE

- (10) *tina -u*
mother-my
my mother

TO'ABA'ITA

- (12) *maka nau*
father I
my father

An indirect possessive construction contains, in addition to the possessed and the possessor, a third element which I will term 'possessive classifier':

FAGANI

- (13) *ga -ku wai*
CLASS-my water
my water

MOTU

- (15) *daika e -na boroma*
who CLASS-his pig
whose pig?

MANAM

- (14) *?usi ne -gu*
loincloth CLASS-my
my loincloth

STANDARD FIJIAN

- (16) *na no -dratou waqa na cauravou*
art CLASS their canoe art young man
(paucal)
the young men's canoe

The elements glossed 'CLASS(IFIER)' in (13)-(16) are usually called 'possessive markers' or 'possessive particles' in the literature dealing with Oceanic languages. As I have argued elsewhere (Lichtenberk 1983a), these forms are really a special kind of classifier. In fact, the term 'possessive classifiers' is used in the descriptions of some Oceanic languages which exhibit a large number of these forms.

Like the more common numeral classifiers, possessive classifiers also classify entities on the basis of some semantic criteria. The difference between the two types of classifier system lies in the nature of the classifying criteria. In numeral classifier systems, entities are classified on the basis of some of their properties (permanent or more or less temporary), such as shape, size, consistency, etc. In the possessive classifier systems of Oceanic languages, the classifying criteria are the types of *relationship* that obtain between two entities, the possessed and the possessor. (It was for this reason that the Oceanic possessive classifiers were termed 'relational' in Lichtenberk 1983a.) The nature of the Oceanic possessive classifier systems will be discussed in more detail in Section 5.

It is necessary to point out that whether a possessive construction is to be considered to be prepositional or indirect is not always obvious from the grammatical descriptions available. A possessive preposition, as the term is applied here, is an element that appears after the possessed and before the possessor and is not a classifier. A classifier is a form that appears in a possessive construction in addition to the possessed and the possessor, and carries a possessive affix or stands in some more or less well-defined semantic opposition to at least one other such form (see Section 5.3 for examples of semantic oppositions between classifiers).

In Kwaio, *a* is a possessive classifier, because even though it does not stand in a semantic opposition to any other such form, it carries possessive suffixes:

- (17) *?ifi a -gu*
house CLASS-my
my house

In Standard Fijian, *kei*, *mei* and *nei* are classifiers, because even though they do not carry possessive suffixes, they stand in well-defined semantic oppositions to each other:

- | | |
|--|--|
| <p>(18) na dalo kei Kolomavu
art taro CLASS K.
<i>Kolomavu's taro (to eat)</i>
(food possessive construction)</p> <p>(20) na isulu nei Bale
art dress CLASS B.
<i>Bale's dress</i>
(general possessive construction)</p> | <p>(19) na yaqona mei Tiuta
art kava CLASS T.
<i>Tiuta's kava (to drink)</i>
(drink possessive construction)</p> |
|--|--|

It may well be that some classifiers that do not carry possessive affixes are grammatically prepositions. What is important for our purposes is that they have a classifying function.

3.3 Simplex and complex possessive constructions

Oceanic possessive constructions can be further classified according to the grammatical status of the possessor element. In a *simplex* construction, the possessor is a possessive affix, either on the possessed or on a classifier:

- | | |
|---|---|
| <p>KOSRAEAN
(21) siyuh-k
<i>belly-my</i>
<i>my belly</i></p> <p>BANONI
(23) ghe -mi numa
CLASS <i>your'(sg) house</i>
<i>your house</i></p> | <p>WAYAN
(22) e -lā
<i>his-leg</i>
<i>his leg</i></p> |
|---|---|

In a *complex* possessive construction, there is a possessor noun phrase separate from the possessed. Three basic subtypes of the complex possessive construction may be distinguished according to the nature of the possessor noun phrase:

- (i) the possessor NP has a nominal head:

- | | |
|---|---|
| <p>TO'ABA'ITA
(24) θaina -na wela lakoo
<i>mother-his child that</i>
<i>that child's mother</i></p> | <p>ARE
(25) wasike a -na bau
<i>woman CLASS-her garden</i>
<i>the woman's garden(s)</i></p> |
|---|---|

- (ii) the possessor NP is an independent personal pronoun:

- | | |
|--|--|
| <p>TO'ABA'ITA
(26) maka nau
<i>father I</i>
<i>my father</i></p> | <p>MANAM
(27) ?ai'o ?ati ne -ŋ
<i>you(sg) canoe CLASS-your(sg)</i>
<i>your canoe</i></p> |
|--|--|

(iii) the possessor NP is a special possessive pronoun (POSS). A number of Oceanic languages have, in addition to the usual set of independent personal pronouns, another set of pronominal forms used only in possessive constructions. The possessive pronouns often bear formal resemblance to the corresponding independent pronouns. In many cases, they appear morphologically complex, but their structure may be far from transparent (see Bradshaw 1982 for a discussion of possessive pronouns in New Guinea languages).

KAIRIRU

- (28) pyal wokyou the first person singular independent
 house I pronoun is kyau
 POSS
 my house
- (29) niu yaqai the third person singular independent
 coconut he pronoun is ei
 POSS
 his coconut grove

EROMANGA

- (30) nimo enugkŭk the second person singular independent
 house you(sg) pronoun is kŭk
 POSS
 your house
- (31) nevag eni the third person singular independent
 food he pronoun is iyi
 POSS
 his food

As will be seen in Section 3.4, type (i) of the complex possessive construction may combine with type (iii).

Prepositional possessive constructions are by definition complex:

HOUAILOU

- (32) virhoε i rhai
 anger of lizard
 the lizard's anger

3.4 Cross-referencing of the possessor

Complex possessive constructions may be further subclassified according to whether the possessor is cross-referenced elsewhere in the construction or not. In (33) and (34), the possessor is not cross-referenced:

TO'ABA'ITA

- (33) nini nau
 knife I
 my knife

HAWAIIAN

- (34) nā kānaka o ke ali'i
 art people CLASS art chief
 the people of the chief

By definition, the possessor is never cross-referenced in prepositional constructions (Section 3.2).

To cross-reference the possessor, the appropriate possessive affix is attached either directly to the possessed (in a direct construction) or to the classifier (in an indirect construction).

BANONI

- (35) garasi-na kara
 glass -its car
 the window glass of the car

KALIAI

- (36) eao ai -awa
 river its-mouth
 the river's mouth

A'ARA

- (37) no -nya mola ihei
 CLASS-his canoe who
 whose canoe?

MANAM

- (38) aine niu ?an -di
 woman coconut CLASS-their
 the women's coconuts

In at least some languages that have possessive pronouns, the possessor may be cross-referenced twice: once by a possessive affix and once by a possessive pronoun:

KAIRIRU

- (39) Nur yaqai qajuo -ny
 N. he cousin-his
 POSS
 Nur's cousin

In a number of languages, the possessor is cross-referenced in certain types of possessive construction but not in others. Whether the possessor is cross-referenced or not may depend on the type of relation between the possessor and the possessed. The possessor may be cross-referenced when the possession is inalienable but not when the possession is alienable (see Section 5.1 for inalienable and alienable possession).

TO'ABA'ITA

- (40) θaina -na θa maeli but (41) luma θa maeli
 mother-his art M. house art M.
 Maeli's mother Maeli's house

Whether the possessor is cross-referenced or not may also depend on the type of the possessor. In Iaai, common-noun possessors are cross-referenced, but proper-noun possessors are not:

- (42) ba -n wanakat but (43) bwə Poou
 head-his child head P.
 the child's head Poou's head

3.5 Complete, partial and construct cross-referencing of the possessor

The cross-referencing of the possessor in a complex possessive construction may be one of three kinds: complete, partial or construct.

(i) Complete cross-referencing: the possessor is cross-referenced both for person and for number.

MOTU

- (44) boroma kwara-na (45) boroma kwara-dia
 pig head -its pig head -their
 the pig's head pigs' heads

A'ARA

- (46) no -gu mola iara
 CLASS-my canoe I
 my canoe

Cross-referencing is always complete when the possessor is other than third person, that is when it is a pronoun (independent or possessive).

(ii) Partial cross-referencing: the possessor is cross-referenced only for person, not for number; the number contrast is neutralised.

TO'ABA'ITA

- (47) θaina -na wela lakoo (48) θaina -na wela ki (**θaina-da...)
mother-his child that mother-his child pl mother-their
that child's mother the children's mother

cf. (49) θaina -da (** signifies ungrammaticality).
mother-their
their mother

(iii) Construct cross-referencing: the possessor is cross-referenced by a special construct suffix which is different from the third person singular and plural possessive suffixes.

MOKILESE

- (50) oadoa-n woall-o cf. (51) oadoa-∅ (with oadoa, the third
name -CONSTR man -that name -his person singular possessive
that man's name his name suffix has a zero form).

In Kosraean, the construct suffix is required only if the possessor is a common noun. If the possessor is a proper noun, the appropriate third person possessive suffix is used:

- (52) sife-n muhtwacn se (53) sifac-l Sepe
head-CONSTR woman this head -his S.
the woman's head Sepe's head

4. TYPOLOGY OF OCEANIC POSSESSIVE CONSTRUCTIONS BASED ON THE ORDER OF THE CONSTITUENTS

Having discussed the basic types of Oceanic possessive constructions from the point of view of the nature of the constituents and the relations between the constituents, we can now turn our attention to the order of the constituents.

In the subsequent discussion, only the independent, i.e. separate, constituents and their ordering will be considered. In other words, the position of possessive affixes will be disregarded. This eliminates from consideration direct simplex constructions, whose structure is POSS'D-POSS'R or POSS'R-POSS'D.

Prepositional possessive constructions exhibit only one order: POSS'D PREP POSS'R. Nothing further will be said about them.

This leaves only three basic types of possessive construction to be considered:

- | | |
|------------------|--------------------------|
| indirect simplex | (POSS'D, CLASS); |
| direct complex | (POSS'D, POSS'R); |
| indirect complex | (POSS'D, POSS'R, CLASS). |

4.1 Indirect simplex possessive constructions

There are two possible orderings of the constituents in indirect simplex constructions:

(i) POSS'D CLASS

MANAM

- (54) pera ?ana -gu
house CLASS-my
my house

(ii) CLASS POSS'D

IAMALELE

- (55) ya -mu vanuga
CLASS-your (sg) house
your house

4.2 Direct complex possessive constructions

Two possible orderings of the constituents exist in direct complex constructions:

(i) POSS=D POSS'R

TO'ABA'ITA

- (56) maka θa Mae
father art M.
Mae's father

(ii) POSS'R POSS'D

AROMA

- (57) Vagi ama -na
V. father-his
Vagi's father

4.3 Indirect complex possessive constructions

Six different orderings of the constituents are theoretically possible in indirect complex constructions, but I have found examples of only five.

(i) POSS'D CLASS POSS'R

KWAIO

- (58) boo a -la Ba?efaka
pig CLASS-his B.
Ba?efaka's pig

(ii) POSS'D POSS'R CLASS

no examples found

(iii) POSS'R CLASS POSS'D

AROMA

- (59) na vavine ge -na vanua
that woman CLASS-her village
that woman's village

(iv) POSS'R POSS'D CLASS

MANAM

- (60) aine niu ?an -di
woman coconut CLASS-their
the women's coconuts

(v) CLASS POSS'D POSS'R

STANDARD FIJIAN

- (61) na no -dratou waqa na cauravou
art CLASS-their canoe art young man
 (paucal)
the young men's canoe

(vi) CLASS POSS'R POSS'D

ULITHIAN

- (62) lema -li Darxos tamaaxoo
 CLASS-CONSTR D. tobacco
Darxos' smoking tobacco

4.4 Harmonic and disharmonic relations between possessive construction types

The three types of possessive construction just discussed can be considered not only by themselves but also in relation to each other in terms of the elements they share. If in a language the order of the elements shared by two possessive construction types is the same in both, the two constructions may be said to be *harmonic* with each other. Thus the following pairs of constructions are harmonic with each other:

POSS'D CLASS (indirect simplex)	and	POSS'R POSS'D CLASS (indirect complex)
POSS'R POSS'D (direct complex)	and	POSS'R CLASS POSS'D (indirect complex).

If the order of the elements shared by two possessive construction types is not the same, the two constructions may be said to be *disharmonic* with each other. The following are examples of constructions disharmonic with each other:

CLASS POSS'D (indirect simplex)	and	POSS'D CLASS POSS'R (indirect complex)
POSS'D POSS'R (direct complex)	and	POSS'R CLASS POSS'D (indirect complex).

The concept of harmonic/disharmonic relations applies only in the comparison of indirect complex (POSS'D, POSS'R, CLASS) constructions with either indirect simplex (POSS'D, CLASS) or direct complex (POSS'D, POSS'R) constructions. Since indirect simplex and direct complex constructions share only one element (POSS'D), there is no relation between them in terms of harmony.

In Manam, both relations are harmonic:

- | | |
|--|--|
| (63) pera ?ana -gu
<i>house CLASS-my</i>
<i>my house</i> | (64) aine niu ?an -di
<i>woman coconut CLASS-their</i>
<i>the women's coconuts</i> |
|--|--|

- (65) boro ta'e -di
pig faeces-their
the pigs' excrements

In Iaaï, direct complex and indirect complex constructions are harmonic with each other, but indirect simplex and indirect complex constructions are disharmonic with each other:

- | | |
|--|---|
| <p>(66) ba -n wanakat
 <i>head-his child</i>
 <i>the child's head</i></p> <p>(68) a -n kəpiŋ
 <i>CLASS-his taro</i>
 <i>his taro</i></p> | <p>(67) wɔɔ a -n θaan
 <i>fish CLASS-his chief</i>
 <i>the chief's fish</i></p> |
|--|---|

In Mokilese, on the other hand, it is the indirect simplex and indirect complex constructions that are harmonic with each other, and direct complex and indirect complex constructions that are disharmonic with each other:

- | | |
|--|--|
| <p>(69) nima -h⁴ penn -ok
 <i>CLASS-his coconut-that</i>
 <i>his coconut</i></p> <p>(71) oadoa-n woall-o
 <i>name -CONSTR man -that</i>
 <i>that man's name</i></p> | <p>(70) nime -n lih -o penn -ok
 <i>CLASS-CONSTR woman-that coconut-that</i>
 <i>that woman's coconuts</i></p> |
|--|--|

This concludes the survey of the basic types of Oceanic possessive constructions in terms of their structures. We are now ready to consider the semantics of the various possessive types.

5. SEMANTICS OF DIRECT, INDIRECT AND PREPOSITIONAL POSSESSIVE CONSTRUCTIONS

5.1 Alienable and inalienable possession

The distinction between alienable and inalienable possession is crucial for the understanding of the use of the possessive constructions in most Oceanic languages. Following Lyons, we may characterise alienable possession as that in which "[the possessed] is contingently associated with [the possessor]", and inalienable possession as that in which "[the possessed] is necessarily associated with [the possessor]" (Lyons 1968:301). Prime examples of inalienable possession are part-whole relations (my head, the roof of the house) and kinship relations (my father). As examples of alienable possession we may give John's car and my pen.

Even though it is subject to language-idiosyncratic exceptions, the following generalisation concerning direct and indirect possessive constructions can be made: *If a language has a distinction between direct and indirect possessive constructions, then direct constructions are used to express inalienable possession, and indirect constructions are used to express alienable possession.* I will return to the problem of exceptions to this generalisation in Sections 5.3 and 7.1. Prepositional possessive constructions exhibit some complications of their own. These will be discussed in Section 5.4.

5.2 Semantics of direct possessive constructions

Direct possessive constructions are normally used to express inalienable possession. Inalienable possession includes the following major categories:

(i) part-whole relations

ROVIANA

(72) bulo -di
heart-their
their hearts

MOKILESE

(73) dame -n warr -o
outrigger-CONSTR canoe-that
that canoe's outrigger

(ii) kinship relations

A'ARA

(74) kma -nya
father-his
his father

MOTU

(75) mero sina -na
boy mother-his
the boy's mother

(iii) spatial relations

IAAI

(76) hpimeka-n
face -his
in front of him

ARE

(77) yove muri -na -i
house behind-its-at
behind the house

The treatment of spatial relations as inalienable possession follows from the fact that the spatial nouns normally refer to parts of a whole, especially body parts, or historically derive from such nouns.

(iv) the concept of 'alone', emphatic '-self'

MANAM

(78) rube -gu u-yalale
alone-my I-go
realis
I went alone/by myself

BABATANA

(79) zira ka jüjini tana-dira
they past do self-their
they did it themselves

IDUNA

(80) tau -na a -na kevakeva gi-bayauma
self-his CLASS-his fish he-caught
he himself caught his fish

Direct possessive constructions may also be used with:

(v) nominalisations, where the possessor encodes the actor (agent, experiencer) or the patient

TO'ABA'ITA

(81) ŋuu -la -na
sing-nom-his
his singing

KOSRAEAN

(82) kahto -iye-n ...
pretty-nom-CONSTR
prettiness of ...

MANAM

(83) udi tanom-a -di
banana plant-nom-their
the planting of the bananas

(vi) physical and mental states or attributes

NAKANAI (84) la vagagari-mu art <i>strong</i> - <i>your</i> (sg) <i>your strength</i>	MANAM (85) tamoata taburi-di man <i>fear</i> - <i>their</i> <i>the men's fear</i>
--	--

The last category – physical and mental states or attributes – is, however, expressed in a number of languages by means of one subtype of indirect possessive construction (see the next section).

5.3 Semantics of indirect possessive constructions

Indirect possessive constructions are normally used to express alienable possession.

One characteristic that Oceanic possessive classifier systems share with numeral classifier systems is that they always have a general, semantically unmarked classifier which is used when none of the more specific possessive constructions is applicable. Some writers refer to this possessive type as 'active manipulative' (Lynch 1982), or as 'dominant' because "the possessor owns or is in physical control of the head noun" (Pawley 1973:158). I prefer the term 'general possessive construction' for two reasons: (1) it is not only in this type of possession that the possessor may own or manipulate the possessed; and (2) this category is clearly semantically neutral. It is used when none of the other, specific categories is appropriate. If a language has only one classifier, the only function of that classifier is to mark alienable possession in opposition to inalienable possession.

The number of possessive classifiers in an Oceanic language ranges from one to well over 20. However, the languages fall into two fairly distinct groups. One has a small number of classifiers – from one to four. The other has a large number of classifiers – upwards of 10, and usually more than 20.

Probably the most widespread possessive classifier system is one with two classifiers: an alimentary one and a general one (but see further below in this section for the use of the term 'alimentary classifier'). The alimentary classifier is used when the possessed is conceived of as an item of food or drink or an object somehow associated with food or drink (food/drink container, implement used in obtaining or preparing food/drink, place where food/drink is obtained). The general classifier is used elsewhere. Manam can be taken as an example of a language of this type. The alimentary classifier is ?ana ~ ?an; the general classifier is ne.

(86) baŋ ?ana -gu taro CLASS- <i>my</i> <i>my taro (to eat)</i>	(87) aine niu ?an -di woman coconut CLASS- <i>their</i> <i>the women's coconuts (to drink)</i>
(88) ?aula ?an -di fishhook CLASS- <i>their</i> <i>their fishhooks</i>	(89) ?usi ne -gu loincloth CLASS- <i>my</i> <i>my loincloth</i>
(90) ?ai?o ?ati ne -ŋ you(sg) canoe CLASS- <i>your</i> (sg) <i>your canoe</i>	(91) natu sioti ne -di child shirt CLASS- <i>their</i> <i>the children's shirts</i>

A few languages have only one possessive classifier. The only function of the classifier is to mark the possession as alienable.

KWAIO

(92) ?ifi a -gu
house CLASS-my
my house

(93) boo a -la Ba?efaka
pig CLASS-his B.
Ba?efaka's pig

If a language has three possessive classifiers, they always are:

(i) food: the possessed is conceived of as an item of food, or an object somehow associated with food (food preparation, storage, etc.) (See below for the use of the term 'food classifier');

(ii) drink: the possessed is conceived of as an item of drink, or an object somehow associated with drink;

(iii) general.

STANDARD FIJIAN

food:

(94) na ke -mudrau kākana
art CLASS your(dual) food
your food

(95) na ke -dratou dalo na qase ni vuli
art CLASS their taro art schoolmaster
(paucal)
taro for the schoolmasters

drink:

(96) na ne -munī tī
art CLASS-your(pl) tea
your tea

general:

(97) na no -mu waqa
art CLASS-your(sg) canoe
your canoe

There are a few languages that exhibit four types of indirect possessive construction. Besides the three discussed above – food, drink and general, they have a category of valued possession.

RAGA

(98) bila -ra boe
CLASS-their pig
their pig(s)

It was said in the preceding section that the uses of direct possessive constructions may include physical and mental states or attributes. However, in a number of languages this category of possession requires not the direct construction but one type of indirect construction where the classifier is of the same form as the alimentary or food classifier. Standard Fijian is one such language.

(99) na ke -na bībī
art CLASS his heavy
his weight

(100) na ke -na levu
art CLASS-his big
his size

The same construction is also used when the referent of the possessor is in some sense a 'sufferer' or 'undergoer':

- (101) na ke -na dali
 art CLASS-*his* rope
his rope, i.e. rope (to be) used on him, such as for binding him or strangling him
- (102) na ke -na itukutuku
 art CLASS-*his* story
his story, i.e. story about him

Pawley (1973) refers to these types of relation between the possessor and the possessed as 'subordinate'. Generally, subordinate possession expresses "actions over which the possessor has no control (where he is the patient, target, or involuntary experiencer)" (Pawley 1973:162). It may also be used with intimate property, such as skirts and loincloths.

The classifier used in subordinate possession in Standard Fijian is *ke* ~ *kei*, which are also the forms of the food classifier; compare (99)-(102) and (95). Pawley (1973 and personal communication) suggests that in Standard Fijian and other languages in which the classifier used in subordinate possession has the same form as the food or alimentary classifier we are dealing with two homophonous classifiers rather than with one with complex semantics. His argument is that since food(/drink) possession and subordinate possession have semantically nothing in common, the two forms are to be considered two separate morphemes.

I prefer the opposite view, according to which in cases of this sort both food(/drink) and subordinate possession require the same classifier. For the time being and for convenience, I will continue to refer to those classifiers as 'alimentary' or 'food'. I will provide arguments supporting my view in Section 6.4.2. Henceforth it should be borne in mind, that in some languages the alimentary or food classifier may also be used to express subordinate possession.

Languages with a large number of possessive classifiers have a proliferation of semantic categories of possession, some of which may be very specific. Kosraean may be taken as a representative of this type of language.

Kosraean has separate possessive classifiers for chewable food, drinkable food, raw/uncooked food, other food, means of transportation, plants, land/shelters, mats, mother and wife, father and husband, and many others, including a general classifier.

- | | |
|---|---|
| <p>(103) sikutuhr okuh -k
 <i>scooter</i> CLASS-<i>my</i>
 <i>my scooter</i>
 (means of transportation)</p> | <p>(104) mos suhnuh-k
 <i>breadfruit</i> CLASS-<i>my</i>
 <i>my breadfruit tree</i> (plant)</p> |
| <p>(105) kuom luh -k
 <i>tray</i> CLASS-<i>my</i>
 <i>my tray</i> (general)</p> | |

There is one lower order subgroup of Oceanic that has no direct possessive construction type: the Polynesian subgroup. Polynesian languages typically have two possessive classifiers, usually of an *o* and a basic form. The two possessive types are often referred to as 'O possession' and 'A possession'. Their use is characterised not so much by the opposition between inalienable and alienable possession but by the concept of control over the initiation of the relation on the part of the possessor. If the possessor does not have control over the initiation of the relationship, O possession is required.

If the possessor does have control, A possession must be used (for more detail on Polynesian possessive constructions see Wilson 1982).

HAWAIIAN

(106) k -o -na lima
art-CLASS-his hand
his hand
(O possession)

(107) k -ā -na keiki
art-CLASS-his child
his child
(A possession)

5.4 Semantics of prepositional possessive constructions

Both alienable and inalienable possession may be expressed by prepositional possessive constructions.

BABATANA

(108) pade ta mamalata
house of uncle
(my) uncle's house

DEHU

(109) la ŋalapa ne la ɔaɔiŋ
art house of art girl
the house of the girl

STANDARD FIJIAN

(110) na yaca ni gone
art name of child
the name of the child

There is a strong tendency, however, for prepositional constructions to express alienable possession.

5.5 Overlap between possessive construction types

A feature commented upon by numerous students of Oceanic languages is the possibility for one and the same noun to appear as the possessed in more than one type of possessive construction, almost always with some difference in meaning. Following Lynch (1973), I will refer to this phenomenon as 'overlap'. The existence of overlap is another characteristic that Oceanic possessive classifier systems share with numeral classifier systems. In the latter, it is often the case that the referent of a noun is classifiable according to more than one parameter, and consequently the noun can appear with different classifiers.

A type of overlap widespread throughout the Oceanic subgroup is one where the choice of the construction depends on whether the possessor is conceived of as playing an active or a passive role in the relationship, that is whether he is an actor or a patient. In languages with a direct-indirect possessive construction contrast, if the possessor plays an active role, it is the general construction that is normally used. If the possessor plays a passive role, that is, if the possessor stands in a subordinate relation to the possessed, one of two other constructions is used: direct in some languages, and alimentary/food in others (Section 5.3).

MANAM

(111) nanari ne -gu
story CLASS-my
my story (that I invented,
told, like, etc.)

(112) nanarita? -a -gu
tell story about-nom-my
my story (about me)

MOTU

- (113) Morea e -na sivarai (114) Morea sivarai-na
 M. CLASS *his story* M. *story -his*
Morea's story (that he told) *Morea's story (about him)*

STANDARD FIJIAN

- (115) na no -na itukutuku (116) na ke -na itukutuku
 art CLASS-*his story* art CLASS-*his story*
his story (that he told) *his story (about him)*
- (117) na no -na dali (118) na ke -na dali
 art CLASS-*his rope* art CLASS-*his rope*
his rope (owned by him) *his rope ((to be) used on him,*
e.g. for binding him or
strangling him)

Further examples of overlap:

MANAM

- (119) baligo -gu (120) baligo ne -gu
grass-skirt-my *grass-skirt CLASS-my*
my grass-skirt (when I *my grass-skirt (when I am not*
am wearing it) *wearing it)*

MARSHALLESE

- (121) kötka-m mā
 CLASS-*your(sg) breadfruit*
your breadfruit tree (which you planted)

The classifier kötka signifies that the possessed was planted by the possessor.

- (122) neji -m mā
 CLASS-*your(sg) breadfruit*
your breadfruit tree (owned by you but not necessarily planted
by you)

neji signifies cherished possession.

- (123) kijö -m mā
 CLASS-*your(sg) breadfruit*
your breadfruit (for eating)

kijö signifies that the possessed is conceived of as an item of food.

TO'ABA'ITA

- (124) θaaθate-mu (125) θaaθate ʔoe
jaw -your(sg) *jaw you(sg)*
your jaw (part of *your jaw (that you have, e.g.*
your body) *a pig's jaw)*

Examples could easily be multiplied. I will return to overlap in Section 7.1.

Having considered in some detail both the forms and the uses of the basic Oceanic possessive types, we can now turn to the other main aim of this paper, which is a reconstruction of the Proto-Oceanic possessive system.

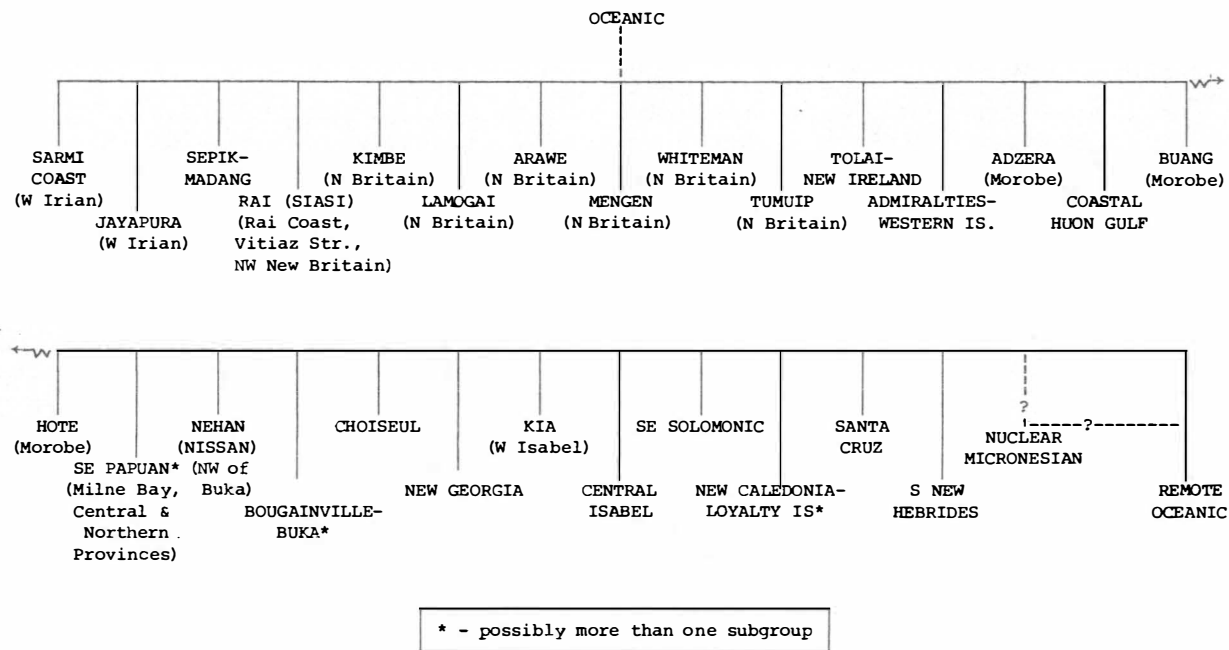


Figure 1: Subgrouping of Oceanic (Pawley 1982a)

6. RECONSTRUCTION OF THE PROTO-OCEANIC SYSTEM OF POSSESSIVE CONSTRUCTIONS

6.1 Methodological preliminaries

In the subsequent discussion, I will assume the subgrouping of Oceanic as shown in Figure 1.

I will follow the standard procedure of comparative linguistics and reconstruct a feature for Proto-Oceanic (POC) if it is found in at least two primary subgroups and its presence there is unlikely to be due to independent parallel developments or to borrowing. To guard against the possibility of borrowing, I will rely, wherever possible, on evidence from geographically distant primary subgroups. As will become apparent, nearly all of the features that I reconstruct for POC are found in many more than just two primary subgroups.

After briefly discussing the recent attempts at reconstructing the POC possessive system, I will proceed to a step-by-step reconstruction using the concepts and categories introduced in the preceding sections.

6.2 Recent attempts at reconstruction

In recent times there have been two major attempts at reconstructing the basic features of the POC system of possessive constructions: Pawley 1973 and Lynch 1982. Interestingly, Pawley and Lynch arrive at entirely different reconstructions.

Using the terminology of the present paper, Pawley's reconstruction can be summarised as follows (only the structures are given at this point, not the forms used in the structures; the forms will be considered in later sections): POC had a set of possessive suffixes used to index and/or cross-reference the possessor. It had direct simplex constructions, whose structure was POSS'D-POSS.SUFF. It also had indirect simplex constructions, whose structure was CLASS-POSS.SUFF POSS'D. Pawley assumes that direct simplex constructions had a classifier of a zero form, although it is not clear how a zero morpheme might have carried possessive suffixes. Direct simplex constructions were used to express inalienable possession; indirect simplex constructions were used to express alienable possession.

Pawley provides no other reconstructions but implies that POC may have had three other constructions of the types that exist in Standard Fijian. One was a prepositional construction with *ni* as the preposition: POSS'D PREP POSS'R. Another was an indirect complex construction: CLASS-POSS.SUFF POSS'D POSS'R. And finally, POC had a direct complex construction, but Pawley does not indicate the order of the constituents.

As will be seen in what follows, the system reconstructed for POC in the present study accords to a large degree, though not entirely, with that postulated by Pawley.

Lynch, in his 1982 study, arrives at conclusions which are radically different from Pawley's. In fact, Lynch does not reconstruct any nominal possessive constructions for POC at all. According to him, the nominal possessive constructions of modern Oceanic languages derive from POC transitive verbal structures. I will return to Lynch's suggestion in Section 7.2.

6.3 Possessive affixes

Did POC have possessive suffixes, possessive prefixes or both? Both suffixes and prefixes are found in more than one primary subgroup. However, there is overwhelming evidence that POC had possessive suffixes and not prefixes. First of all, none of the modern languages has only prefixes. That is, every language that has possessive affixes has suffixes and a few also have prefixes. Furthermore, the type and/or the use of the prefixes is always highly restricted.

In Wayan, the prefixes are used only in direct possessive constructions but (with a few exceptions) only when the possessed is not a kinship term. Elsewhere in direct and everywhere in indirect constructions, the suffixes are required. Basically the same is true of all the other Western Fijian communalects (Geraghty 1983).

In Kaliai, prefixes exist only in the third person singular: ai- and e-. The former is used in direct possessive constructions and one type of indirect construction. The latter is used in the other type of indirect construction. Besides the two prefixes, there is also a third person singular suffix -ai also used in direct constructions. (The factors governing the distribution of the two ai affixes are not stated in the description of the language.)

In Nguna, a few kinship terms take prefixes instead of suffixes, and a few others require both. The prefixes are identical in form with the corresponding suffixes.

Secondly, when one considers the forms of the prefixes, one finds no evidence for reconstructing a set of prefixes for POC. The only instance of formal similarity between forms found in different primary subgroups that I have come across is the third person singular prefix, which is e- in Western Fijian communalects and in Kaliai. Given that this is the only case and that the prefix consists of a single segment, the identity is best attributed to chance.

There are formal correspondences between some of the prefixes of the Western Fijian languages and those of Nguna:

	Western Fijian	Nguna
1sg	qu-	gu-
2pl	m̄-	mu-
3pl	dra-	ta-

However, Western Fijian and Nguna belong in the same subgroup, and furthermore, the prefixes are identical with the corresponding suffixes in these two languages.

When one turns to possessive suffixes, one finds a veritable embarrassment of riches. Not only are possessive suffixes, or at least traces thereof, exhibiting systematic formal correspondences found throughout Oceanic; in some cases one is even forced to reconstruct more than one proto-form. The following is the set of possessive suffixes reconstructible for POC. In most cases, the supporting evidence cited is only a fraction of that available.

- 1st singular: *-ŋku; Standard Fijian, Roviana, Tolai -qu, Manam, Suau -gu.
 2nd singular: *-mu; To'aba'ita, Dobu, Babatana, Ulithian, Kaliai -mu.
 3rd singular: *-ña; A'ara -nya, Bonkovia -ña, Wayan -ya ~ -a,
 Kairiru -ny, Kaniet -ñ.
 *-na; Luqa, Kubokota, Bonkovia -na, Wogeo, Loniua -n.

All the languages in the latter group have separate reflexes of POC *n and *ñ. Note that Bonkovia apparently has both -ña and -na (Blust 1978a:150, note 46, citing material collected by D. Tryon). For discussions of the conflicting reflexes of the third person singular possessive suffix see Blust 1978a and Lichtenberk 1979.

- 1st plural exclusive: *-mami; Arosi, Roviana, Maewo, Babatana -mami, Banoni -mam.
 *-mai; Balawaia, Kaliai, Paama, Tami, Kia -mai.
 *-ma; Dobu, Gedaged, Lakon, Manam, Bareke -ma.
 *-mi; A'ara, Maewo, Nemi, Nissan -mi, Sonsorol-Tobi -mi (√-memi √-memi)⁵.

Maewo has both -mami and -mi. Sonsorol-Tobi also has forms which point to POC *-mami and *-mi.

- 1st plural inclusive: *-nta; Manam, Bugotu, Standard Fijian -da, Tigak, Banoni -ra.
 2nd plural: *-miu; Mota, Nissan, Suau, Vaturanga, Luqa -miu.
 *-mi; Are, Banoni, Gedaged, Marovo, Volow -mi.
 *-mu; Laghu, Nguna, Efate, Lakon -mu.
 3rd plural: *-ndia; Bugotu, Roviana -dia, Balawaia, Mono-Alu -ria, Gilbertese -ia (??).
 *-ndi; A'ara, Dobu, Roviana -di, Banoni, Tigak -ri.
 *-nda; Standard Fijian -dra, To'aba'ita -da, Eromanga -nda, Nenemas -la, Iaa'i -ḍa √ -ra.

Roviana has both -dia and -di.

To summarise:

POC possessive suffixes

1sg	-ŋku
2sg	-mu
3sg	-ña, -na
1pl excl	-mami, -mai, -ma, -mi
1pl incl	-nta
2pl	-miu, -mi, -mu
3pl	-ndia, -ndi, -nda

Systems of possessive markers similar to the one reconstructed here for POC have been reconstructed for Proto-Austronesian (PAN) by Dyen (1974) and Blust (1977) and for Proto-Eastern Oceanic (PEO) by Pawley (1972).

PAN possessive suffixes (Dyen 1974)

1sg	-ku? ~ ŋku?
2sg	-Xu?, -mu? ~ -m
3sg	-ña?
1pl excl	-mami?, -mai?, -ma?, (?)-mi?, -mia?, -kai, -mæŋ, -mamu? (?)
1pl incl	-ta? ~ -nta?
2pl	-mi?, -mu?, -miu?
3pl	-Da? ((?) -nDa?), (?)-Dyæ?/-Dya?, -Di

PAN genitive pronouns (Blust 1977)

1sg	i-ku/ni-ku	
2sg	i-Su/ni-Su	(later on, in Proto-Malayo-Polynesian the plural forms i-mu/ni-mu were used as polite forms for the singular)
3sg	i-a/ni-a	
1pl excl	i-mi/ni-mi	
1pl incl	i-ta/ni-ta	
2pl	i-mu/ni-mu	(in Proto-Malayo-Polynesian also iSu/niSu)
3pl	i-Da/ni-Da	

Note that in Blust's reconstruction, the pronominal forms are not suffixes but genitive pronouns preceded by prepositions *i or *ni.

PEO possessive suffixes (Pawley 1972)

1sg	-(ŋ)ku
2sg	-mu
3sg	-ña
1pl excl	-mami
1pl incl	-(n)ta
2pl	-m(i)u
3pl	-nda

Contrary to Pawley (1972), I have found no evidence for indeterminacy between *k and *ŋk in the first person singular and between *t and *nt in the first person plural inclusive in PEO. The problem of plain versus so-called prenasalised stops in those forms is, however, part of the phenomenon of oral-nasal crossover in Austronesian languages, and it would not be surprising to find conflicting evidence in the reflexes of the two suffixes.

Interestingly, with the exception of the third person singular, in all the other cases where multiple reconstructions are necessary, there is one form which in a sense contains the others. That is, those forms may be viewed as different reductions of the longest form:⁶

1pl excl	-mami and -mai, -ma, -mi
2pl	-miu and -mi, -mu
3pl	-ndia and -ndi, -nda.

One could make a strong case for positing only the longest forms for POC and for assuming the existence of the shorter forms in different languages to be due to independent identical reductions. Another factor which may be responsible for the multiplicity of forms is analogical realignment of the possessive suffixes with the corresponding forms in the other pronominal sets: independent, subject-marking and object-marking (see Simons 1980 for a discussion of analogical influences between pronominal sets in Malaitan languages).⁷ The multiple reconstructions given here are not to be interpreted as alternants in POC but as an indication of the existence of multiple sets of cognates in the Oceanic subgroup.

Evidence from present-day languages indicates that the dual and the paucal possessive suffixes were derived in POC from the plural forms by means of dual and paucal suffixes respectively.

The existence of a special construct possessive suffix appears restricted to Micronesian languages and as such is clearly an innovation.

6.4 POC possessive construction types

In this section, only the possessive construction types as such are reconstructed. The nature and form of some of their constituents and the order of the constituents will be considered subsequently.

Comparative evidence makes it abundantly clear that the following possessive construction types are to be reconstructed for POC:

- (i) direct simplex
- (ii) direct complex
- (iii) indirect simplex
- (iv) indirect complex

All four types are widely distributed within Oceanic (see Section 6.7).

6.4.1 Simplex and complex constructions

The existence of possessive suffixes in POC having been established in Section 6.3, it follows that POC direct simplex constructions were of the form POSS'D-POSS.SUFF.

POC also had indirect simplex constructions, where the possessive suffixes were attached to a classifier.

As far as complex possessive constructions are concerned, recall that there are three types of noun phrase that may function as the possessor in present-day languages: a noun phrase with a nominal head, an independent personal pronoun, and a possessive pronoun. That POC had complex constructions where the possessor was a noun phrase with a nominal head may be safely concluded from the presence of this type of construction in every modern Oceanic language for which relevant data are available. What about the other two categories? Let us consider possessive pronouns first.

Possessive pronouns are found in languages that belong to different first-order subgroups of Oceanic, for example Kairiru, Yabem, Sesake. Possessive pronouns usually appear to be morphemically complex even though their composition is often far from obvious. Usually they seem to consist of a form reminiscent of the corresponding possessive suffix or the independent pronoun preceded by another element, which in some cases at least, may have originally been a possessive classifier. Consider the Kairiru and the Hote possessive pronouns given below. In Kairiru, they bear resemblance to the independent pronouns. In Hote, on the other hand, they exhibit a formal connection with the possessive suffixes.

KAIRIRU	possessive pronouns	independent pronouns
1sg	wokyau	kyau
2sg	yieqayieq	yieq
3sg	yaqai	ei
1pl excl	taqait	qait
1pl incl	tamoit	taqam
2pl	maqam	qam
3pl	rraqarri	rri

HOTE	possessive pronouns	possessive suffixes
1sg	yanəŋ ~ yenaŋ	-ŋ
2sg	anim	-m
3sg	aneŋ	-ŋ
1pl excl	iniŋ ~ aniŋ	-ŋ
1pl incl	iniŋ ~ aniŋ	-ŋ
2pl	unim	-m
3pl	iniŋ	-ŋ

(There is a great deal of syncretism in the Hote possessive suffixes. In fact, the only distinction is between the second person and the rest. There is no number distinction in the possessive suffixes. In the possessive pronouns, there is no exclusive-inclusive distinction. All of these distinctions are, however, found in the independent pronoun set.)

The only similarity that possessive pronouns exhibit across subgroups is that between those parts of them that derive historically from possessive suffixes or from independent pronouns. One must conclude that the existence of possessive pronouns in different subgroups is due to independent developments. Consequently, no possessive pronouns are to be reconstructed for POC.

The question of whether POC allowed independent personal pronouns as the possessor in complex constructions is a different problem altogether. Obviously a set of independent pronouns is to be reconstructed for POC. As I will suggest further below, in POC complex constructions the possessor was cross-referenced elsewhere in the construction by means of a possessive suffix. The question then is: Did POC have complex possessive constructions in which the possessor was an independent pronoun even though there was a corresponding possessive suffix elsewhere in the construction?

Among present-day languages, there are many that do allow this type of construction, but there are many others that do not.

MANAM	TO'ABA'ITA
(126) ŋa ruaŋa -gu	(127) θaina -na **nia
<i>I friend-my</i>	<i>mother-his he</i>
<i>my friends</i>	<i>his mother</i>

The possibility of co-occurrence of an independent pronoun with a possessive suffix has been believed by some to be characteristic of New Guinea Oceanic languages (e.g. Milke 1965, Pawley 1978 among the more recent ones). However, this type of construction is not restricted to New Guinea:

BABATANA	NISSAN
(128) tama -mu re	(129) ingo tar lima-g
<i>father-your you(sg)</i>	<i>I art arm -my</i>
<i>your father</i>	<i>my arm</i>
ULAWA	MAROVO
(130) inau maraa-ku	(131) isu -da hita
<i>I self -my</i>	<i>nose-our we (incl)</i>
<i>I myself</i>	<i>our noses</i>

Such constructions appear to be used in various discourse processes, such as focusing and contrast. The use of independent pronouns in addition to possessive affixes in possessive constructions is thus reminiscent of their use in subject and object positions together with the appropriate — usually clitic or affixal — subject and object markers. Compare the following two examples from Manam:

- (132) ?ai?o ?ati ne -ŋ
 you(sg) canoe CLASS-your(sg)
 your canoe
- (133) ?ai?o ?alea naŋatai-o ?u -pura?
 you(sg) month which -in you(sg) realis-arrive
 in which month did you arrive?

It is evident that the use of independent pronouns in possessive constructions is part of a larger pattern and that it must be considered in conjunction with their use in other types of construction. In order to determine whether POC allowed independent pronouns to co-occur with possessive suffixes, it will be necessary to have some idea about the processes of focusing, topicalisation, contrast, etc. in POC. If POC did have a way of foregrounding pronominal elements, it is highly likely that independent pronouns could co-occur with possessive suffixes in complex possessive constructions.

6.4.2 Indirect constructions

Pawley (1973) has reconstructed three possessive classifiers for POC, including a pair of homophones: *ka alimentary/food, *ka subordinate, and *na general (called 'dominant' by Pawley). He also suggests that POC may have had a drink classifier *ma but says that the evidence is rather weak.

It is clear that a classifier for food possession must be reconstructed for POC. For the time being, I will call it a 'food' classifier.

Food classifier: *ka; Tami ka, Vaturanga, Marovo ha, Dobu 'a, Mota ga.

A general classifier also must be reconstructed for POC. Here, however, the reconstruction is much more difficult owing to irregular vowel correspondences. Assuming that the following forms are cognate, all five of the usual Oceanic vowels are found: Bugotu ni, Manam ne, Kubokota na, Standard Fijian no, Duke of York Island nu.

The form nu is found only in the language of Duke of York Island and can therefore be left out of consideration. ni occurs in several South-East Solomonic languages and in Gitua. no is common in Remote Oceanic. It is also found in Kia, Central Isabel languages, the Reefs language of Santa Cruz and in Nissan (in alternation with na). ne is found in Manam, Tami, a number of Remote Oceanic languages, usually in alternation with other forms, and also in some of the Santa Cruz languages. na occurs in Remote Oceanic, Kubokota, Nissan (in alternation with no), Malo (in alternation with ne), Babatana (as part of the singular possessive pronouns), Roviana (as part of the third person singular and the first person plural inclusive possessive pronouns), and possibly in Gedaged (ina).

Pawley opts for reconstructing *na as allowing the most economical explanation of the forms showing other vowels, although he considers only the ne and no variants. He suggests that the ne and no (and by extension also the ni and nu) forms arise from various kinds of assimilation of *na to the high vowels of many of the possessive suffixes. The assimilated form was later on generalised to appear before all possessive suffixes. My investigation has uncovered no evidence that would contradict Pawley's reconstruction, and consequently I have retained it.

One more point needs to be made in this respect. It has been noted by various investigators that in many languages the alimentary/food classifier bears formal resemblance to the word for *eat* and/or *food*; compare also POC *ka *food classifier* and *kani *eat*. One characteristic of classifiers is that they typically derive historically from lexical morphemes, usually nouns. This is true both of numeral classifiers and of possessive classifiers of Oceanic languages. In Oceanic languages, the classifier for means of transportation normally derives from the word for *canoe*; the classifier for shelters usually comes from the word for *house*; etc. It is thus quite likely that the POC food classifier derives ultimately from the word for *food*, or *eat* via a nominalisation.

One would expect the general classifier to derive from a semantically-neutral lexical item, perhaps *thing*. Finding a plausible source for the general classifier might help us to determine the form of the classifier in POC. My less-than-thorough search for a possible source of the general classifier has turned up the following forms for *thing*: Lenakel nar, Atchin na, also na-na, Amari nam (also *food*). Whether the forms are cognate is impossible to tell at this point. Interestingly, however, they all have a as their vowel. If a word for *thing* should be reconstructible for POC on the basis of these forms, it would provide strong support for reconstructing the general classifier as *na. (Note, however, that Reid (1983) argues that the form of the POC general classifier was *ano, reflecting Proto-Malayo-Polynesian *anu *thing*.)

Let us now consider the possibility of POC having a drink classifier. Pawley tentatively reconstructs *ma as a drink classifier on the basis of data from a number of Remote Oceanic languages, Lenakel of the Southern New Hebrides, and Vanikolo and Utupua of Santa Cruz. However, he points out that Lenakel and the Santa Cruz languages are geographically not very distant from the Northern New Hebrides and the Banks Islands, where many of the Remote Oceanic languages with drink classifiers are located. Consequently, the possibility of borrowing cannot be excluded. According to Pawley, there are no reflexes of *ma anywhere in New Guinea or north-west Melanesia.

There do, however, exist languages in north-west Melanesia with possessive classifiers that point to POC *ma. According to Codrington (1885), the language of Duke of York Island (between New Britain and New Ireland) has a drink classifier ma. In his discussion of the possessive classifiers of Duke of York Island, Codrington says about ma: "With things to drink ma is used as in Mota" (Codrington 1885:568). Unfortunately, he does not give an example. Ray (1919), when discussing the (Standard) Fijian drink classifier me, says: "me is the Fijian equivalent of ma used as a possessive in the New Hebrides, Banks Islands, and Bismarck Archipelago" (p.354). Again unfortunately, Ray provides no detail about ma in the Bismarck Archipelago (in fact, he may have based his statement on Codrington 1885).

Recently, new evidence has come to hand concerning the existence of reflexes of POC *ma in north-west Melanesia. According to Johnston (1983), Tomoip of north-east New Britain has a drink classifier alma, which may be bimorphemic; cf. al *kin possession* and ma *inalienable possession*.

Secondly, M. Ross has informed me that the language of Tabar (off New Ireland) has a possessive classifier of the form ma. Note, however, that the classifier marks not just drink possession but alimentary possession in general. As in many other Oceanic languages, in Tabar the original food-drink distinction has been replaced by a more general alimentary category. What is unusual about the Tabar case is that it is the original drink not the food classifier that today marks alimentary possession.

Drink classifiers are also found in other languages of Santa Cruz besides Vanikolo and Utupua, cited by Pawley. Wurm (1972) gives the following forms: Reefs numwā, Nemboi mu, Nooli mu. It must be pointed out, however, that Wurm does not consider these languages to be Austronesian, but rather Papuan with heavy Austronesian influence. For a debate concerning the genetic status of these languages see Wurm 1978 and Lincoln 1978.

A drink classifier then appears reconstructible for POC.

Drink classifier: *ma; Mota, Vanikolo, Tabar ma, Standard Fijian me, Lenakel n+mw.

The classifier may be ultimately related to the POC word for *drink* *inum (see especially the Lenakel and the Reefs forms).

The question to be considered next is whether POC had a separate classifier for subordinate possession. Recall that Pawley (1973) has reconstructed a subordinate classifier *ka, homophonous with the food classifier. His reason for positing two classifiers is semantic. Since food possession and subordinate possession have nothing in common (apart from both being possessive relations), the two *ka forms are distinct morphemes. The same argument applies in the case of those present-day languages in which subordinate possession requires a classifier that is formally identical with the alimentary/food one.

I want to argue, however, that in POC there was only one *ka classifier, which was used to express both food possession and (certain kinds of) subordinate possession. My argument is as follows: To my knowledge, in every language that has a direct-indirect possessive type contrast, subordinate possession is expressed in one of two ways: (i) by means of the direct construction, or (ii) by means of an indirect construction where the classifier is formally identical to that used in the alimentary/food construction. There is no language that has a formally-unique subordinate classifier. Furthermore, there are no languages with an indirect subordinate construction but no alimentary/food construction. On the other hand, there are many languages that have an alimentary/food classifier but no subordinate classifier. If one assumes that POC had two *ka classifiers, then one has no explanation for the non-random loss of one of the two forms. Under this assumption, if a language has lost one of the two classifiers, it is always the subordinate one, never the food one that is lost.

What I want to suggest is that POC had only one *ka classifier, which was used to express food possessive relations. As far as subordinate possession is concerned, present-day evidence indicates that both the direct and the *ka indirect constructions were used. Perhaps one was used to express certain kinds of subordinate possession, and the other was used for other kinds of subordinate possession. Through time, languages may have come to prefer one type of construction over the other. Pawley himself says that in Standard Fijian the use of the subordinate classifier has been extended beyond its presumed POC range: it is used with most natural attributes, whereas other languages use the direct construction. On the other hand, at least one kind of subordinate possession requires the direct construction in Standard Fijian. Compare the following two examples:

(134) na yaca-na
 art name-his
 his name (which he is
 called by; i.e. the name
 bestowed on him)

(135) na no -na yaca
 art CLASS-his name
 his name (i.e. the name which
 he has bestowed on somebody)

The relation between (134) and (135) is parallel to that between (136) and (137):

- | | |
|---|--|
| (136) na ke -na dali
art CLASS- <i>his</i> rope
<i>his rope ((to be) used on
 him, e.g. for binding him
 or strangling him)</i> | (137) na no -na dali
art CLASS- <i>his</i> rope
<i>his rope (owned by him)</i> |
|---|--|

From now on, I will refer to the *ka classifier as food/subordinate. Similarly, if in a modern language alimentary/food possession and subordinate possession are formally identical, I will consider the language to have one alimentary/subordinate or food/subordinate classifier, not two homophonous ones.⁸

Pawley also hints that POC may have had a valued possession classifier. Valued possession classifiers are found in languages of Northern Vanuatu and also in Micronesian languages. As Capell (1943) points out, Nada of Milne Bay Province of Papua New Guinea has a form bula "referring to chattels and animal property" (p.229):

- (138) to -gu bula mwila
 CLASS-*my* banana
my banana

The valuable possession classifiers of the Vanuatu languages appear cognate with the Nada form, for example: Gog, Tangoa bula, Mota, Kiai, Leon pula. The forms of the classifiers in the Micronesian languages (not considered by Pawley and Capell) are not cognate and are clearly independent developments.

Concerning the Nada form bula, Capell says that it does not take possessive suffixes. In fact, as (138) demonstrates, bula co-occurs with a classifier. In this respect, it is of course quite different from the valued possession classifiers of the Vanuatu languages. Even though the Nada and the Vanuatu forms may be cognate, the proto-form is unlikely to have functioned as a classifier in POC. Rather it probably was a lexical morpheme ('valuable property?') which has developed into a classifier in some of the languages of Vanuatu.

The conclusion then is that POC had three possessive classifiers: *ka food/subordinate, *ma drink and *na general. Interestingly, all three have a as their vowel. Whether this similarity is accidental or systematic is a question that cannot be answered here.

6.5 Prepositional constructions

There are two prepositions that are found in possessive constructions in a number of languages. Both have to be reconstructed for POC although not, as I will suggest below, with a possession-marking function. The two prepositions are *ni and *qi. Recall also that *ni and *i have been reconstructed as genitive prepositions for PAN (Blust 1977; see Section 6.3). Reflexes of *ni are used in possessive constructions in, for example, Standard Fijian and Nguna. Reflexes of *qi are used with the same function in, for instance, Houailou and Tolai.

Reflexes of both prepositions are, however, more commonly used in NP constructions other than possessive. They are often used in generic NP constructions. In a generic construction, the attribute NP is non-specific, and the construction as a whole indicates a *type* of entity:

ULAWA

- (148) nima ni mwane
house of man
holy house (house for
males) (i.e. a type of
house)
- (140) nima inau
house I
my house

KWARA'AE

- (141) 'ae 'i wae
leg of man
the leg (of man),
(i.e. human leg)
- cf. (142) 'ae-na wae
leg-his man
the leg of a particular man

Reflexes of *ni can also serve to express purpose:

STANDARD FIJIAN

- (143) na bilo ni tī
art cup for tea
tea-cup
- (144) na bilo tī
art cup tea
cup of tea

TO'ABA'ITA

- (145) kafo ni ku'u -laa
water for drink-nom
water for drinking

Reflexes of *qi may express purpose or quantity:

KWARA'AE

- (146) makelea 'i gwata
enclosure for swine
an enclosure for swine
- (147) binu 'i kafo
cup of water
a cup of water

The functions of *qi and *ni have been the subject of a detailed analysis by Hooper (in this volume). Hooper concludes that in POC both prepositions were used in NP constructions other than possessive. She suggests that *qi was used with nouns that could appear in direct possessive constructions (in her words, 'inalienable'), while *ni was used with nouns that normally appeared in indirect possessive constructions ('alienable'). Whatever the distinction between *qi and *ni may have been, neither was used in possessive constructions as these are defined here.

Pawley (1973) also reconstructs a preposition *ta for POC. However, even though ta is used in possessive constructions in a few languages (Babatana, Roviana), the usual function of the reflexes of *ta is to mark spatial relations, most commonly location or place of origin:

MAEWO

- (148) laqana ta Maewo
language of M.
language of Maewo

As Pawley says, *ta most likely marked these spatial relations in POC and the possessive uses of its reflexes in some languages are a later innovation.

We can conclude then that no prepositional possessive constructions are to be reconstructed for POC.

6.6 Cross-referencing of the possessor

Three types of cross-referencing of the possessor in complex possessive constructions are found in present-day Oceanic languages: complete, partial and construct. Construct cross-referencing appears restricted to Micronesian languages and is clearly a later innovation.

Of the other two types, complete cross-referencing is clearly the more widespread and consequently is to be reconstructed for POC.

6.7 The order of the constituents

Thus far we have concluded that POC had both direct and indirect simplex and complex possessive constructions. The direct simplex construction had the structure POSS'D-POSS.SUFF. The problem to be considered next is the order of the constituents in the other three types.

6.7.1 Indirect simplex constructions

Indirect simplex constructions consist of a possessed and a classifier. Both orders — CLASS POSS'D and POSS'D CLASS — occur, and both are common. There is, however, strong evidence that suggest that CLASS POSS'D is the original order. In terms of subgroups, it is more widespread than the other order. Available evidence indicates that it is the only order in the New Georgia, Choiseul, and Bougainville-Buka subgroups. With just a few exceptions, it is the only order in all of the subgroups located in New Guinea. It is clearly the predominant order in the Tolai-New Ireland subgroup.

The other order — POSS'D CLASS — is exclusive only in the Santa Cruz subgroup. Elsewhere, the two orders co-exist, but CLASS POSS'D is clearly the more common one.

The evidence then points to reconstructing the order CLASS POSS'D for indirect simplex possessive constructions in POC. This is also the order reconstructed by Pawley (1973).

6.7.2 Direct complex constructions

Direct complex constructions consist of a possessed and a separate possessor. Comparative evidence is overwhelmingly in favour of reconstructing POSS'D POSS'R as the order in POC. The reverse order — POSS'R POSS'D — is, with a few exceptions, restricted to the New Guinea area and Micronesia. Although in New Guinea it is found in several primary subgroups of Oceanic, it is best considered an areal phenomenon which shows a strong affinity with SOV basic word order so common in the New Guinea area.

Pawley (1973) does not provide any reconstructions of complex possessive constructions but implies that those with animate possessors were extensions of the corresponding simplex constructions. For inanimate possessors, he postulates a prepositional construction POSS'D ni POSS'R.

6.7.3 Indirect complex constructions

Indirect complex constructions consist of three elements: a possessed, a possessor and a classifier. Six orderings of the elements are theoretically possible:

- | | |
|---------------------------|--------------------------|
| (i) POSS'D CLASS POSS'R | (ii) POSS'D POSS'R CLASS |
| (iii) POSS'R CLASS POSS'D | (iv) POSS'D POSS'D CLASS |
| (v) CLASS POSS'D POSS'R | (vi) CLASS POSS'R POSS'D |

Order (ii) seems not to occur and can thus be excluded from consideration.

Orders (iii), (iv) and (vi), in which the possessor precedes the possessed, are found almost exclusively in the New Guinea area and in Micronesia. As pointed out in Section 6.7.2, the POSS'R POSS'D order in the New Guinea languages is an areal phenomenon. The three orders can thus be eliminated as candidates for reconstruction.

This leaves us with two orders: (i) POSS'D CLASS POSS'R, and (v) CLASS POSS'D POSS'R. Both orders are more or less equally distributed within Oceanic and so this kind of evidence cannot serve as a guide for reconstruction. There are, however, two kinds of indirect evidence that can be used, and both point in the same direction.

Firstly, in the languages in which the possessor precedes the possessed, i.e. types (iii), (iv) and (vi), with one exception the classifier always precedes the possessed. Manam is the only language for which I have data that is of type (iv) POSS'R POSS'D CLASS. All the others are either of type (iii) POSS'R CLASS POSS'D, or of type (vi) CLASS POSS'R POSS'D. Assuming that placing the possessor before the possessed in the New Guinea and the Micronesian languages is a later development, languages of types (iii) and (vi) can be taken as evidence for reconstructing a POC order in complex constructions in which the classifier precedes the possessed.

The other type of indirect evidence is the concept of harmonic/disharmonic relations between possessive construction types. Two types of possessive construction are said to be harmonic with each other if the order of the elements shared by the two types is the same in both. If the order of the shared elements is not the same, the constructions are said to be disharmonic with each other (Section 4.4).

Available evidence reveals a strong tendency for languages to have harmonic rather than disharmonic relations between their possessive types. Now, we have already reconstructed CLASS POSS'D as the order in POC indirect simplex constructions. Assuming that, like in the modern languages, the relation between the POC indirect simplex and indirect complex constructions was more likely to be harmonic than disharmonic, order (v) CLASS POSS'D POSS'R is indicated for POC. Notice that this order is also harmonic with the direct complex type, which has been reconstructed as POSS'D POSS'R.

The order CLASS POSS'D POSS'R reconstructed here for indirect complex construction was also implied for POC by Pawley (1973).

6.8 Semantics of the direct and the indirect constructions

It is abundantly clear from comparative evidence that the POC direct constructions expressed inalienable possession (even though as in many present-day languages there may have been individual exceptions). The category of

inalienable possession included part-whole relations, kinship relations, spatial relations and the concepts corresponding to the meanings of English emphatic *-self* and *alone*. It probably included at least some nominalisations. Finally some kinds of subordinate possession most likely required the direct constructions.

As far as the indirect constructions are concerned, POC had three possessive classifiers:

A food/subordinate classifier **ka*, used: (i) when the possessed was conceived of as an item of food or an object somehow associated with food (food preparation, storage, etc.); (ii) to express some kinds of subordinate possession. More research is required to determine which kinds of subordinate possession required the direct construction and which required the classifier **ka*.

A drink classifier **ma*, used when the possessed was conceived of as an item of drink or an object somehow associated with drink.

A general classifier **na*, used whenever none of the other three, more specific types was appropriate.

6.9 Summary of the reconstructions

The following is the POC system of possessive constructions as reconstructed here:

Direct simplex:	POSS'D-POSS.SUFF
Direct complex:	POSS'D-POSS.SUFF POSS'R
Indirect simplex:	CLASS-POSS.SUFF POSS'D
Indirect complex:	CLASS-POSS.SUFF POSS'D POSS'R

Possessive suffixes:

1sg	-ŋku	1pl excl	-mami, -mai, -ma, -mi
		1pl incl	-nta
2sg	-mu	2pl	-miu, -mi, -mu
3sg	-ña, -na	3pl	-ndia, -ndi, -nda

Type of possessor in complex constructions: (i) NP with a nominal head; and possibly (ii) independent personal pronoun.

Type of cross-referencing of the possessor in complex constructions: complete.

Direct constructions: used to express inalienable possession and some kinds of subordinate possession.

Indirect constructions: used to express alienable possession.

Classifiers used in indirect constructions: (i) food/subordinate **ka*; (ii) drink **ma*; (iii) general **na*.

7. SOME RESIDUAL ISSUES

In this concluding section, I want briefly to consider three further questions relevant to the Oceanic possessive systems.

7.1 Gender systems or not?

In the earlier Oceanic descriptive literature (e.g. Codrington 1885, Ray 1926), it was often said that a language with more than one type of possessive construction had a gender or noun-class system. Class membership was said to be determined by the type of possessive construction in which a noun appeared as the possessed. Thus a language might have an inalienable class, a food class, a drink class and a general class. Even some more recent grammars speak of noun classes or genders: Counts 1969 for Kaliai and Milner 1972 for Standard Fijian. Often, however, it is noted that a noun may belong in more than one class because it can occur in more than one type of possessive construction.

In recent times, a different view of Oceanic possessive constructions has been adopted by many students of Oceanic languages. According to this view, the type of possessive construction required is determined not by the class of the possessed but by the nature of the semantic relation between the possessor and the possessed (e.g. Schütz and Nawadra 1972 for Standard Fijian, Pawley 1973 for Standard Fijian, Wilson 1976a for Hawaiian, Wilson 1976b for Hawaiian and Standard Fijian, Lichtenberk 1983b for Manam, Wilson 1982 for Polynesian languages, Lynch 1973, 1982 for Oceanic in general). On this view, the ability of a noun to appear in more than one type of possessive construction is due to the fact that its referent can stand in more than one kind of relationship with the referent of the possessor. A coconut may be viewed as an item of drink, an item of food, or neither (e.g. as the object of a commercial transaction). Lynch (1982:143) even goes so far as to proclaim:

The notion that, in Oceanic (OC) languages, the nature of the possessive construction was determined by the 'gender' of the possessed nominal was finally dispelled in 1973 in Lynch (1973) and Pawley (1973).

The trouble with this sweeping generalisation is that there are numerous exceptions to it. There are a number of languages in which the type of possessive construction is not predictable from the nature of the relationship between the possessor and the possessed. Below are given a few examples where the type of construction required by the possessed is not predictable and must be specified as an idiosyncratic property of the noun.

As noted in previous sections, kinship terms are normally treated as inalienables. In Houailou, however, only some kinship terms are treated as inalienables, while others are treated like alienables. There appears to be no way to predict on semantic grounds which type of construction a given noun requires.

inalienable	alienable
<i>mother</i>	<i>father</i>
<i>spouse</i>	<i>grandfather</i>
<i>sibling</i>	<i>grandmother</i>
<i>child</i>	<i>son</i>
	<i>maternal uncle</i>

In Kairiru, most body parts are treated as inalienables, but many are treated like alienables:

inalienable	alienable
<i>elbow</i>	<i>knee-cap</i>
<i>hand/arm</i>	<i>finger/toe</i>
<i>ear</i>	<i>hair</i>
<i>skin</i>	<i>flesh</i>

There are even pairs of synonyms in Kairiru one member of which is treated as inalienable and the other like alienable:

	inalienable	alienable
<i>chin</i>	ngapoye	ngies
<i>head</i>	ipwo	qarai

A complication of a different sort exists in To'aba'ita. In To'aba'ita, all body-part nouns are treated as inalienables unless they are modified by an adjective or a numeral, in which case they are treated like alienables:

- | | | | | |
|-------|--------------------------|-----|-------|---------------------------|
| (149) | ʔaba-mu | but | (150) | ʔaba mauli ʔoe |
| | <i>hand-your</i> (sg) | | | <i>hand left you</i> (sg) |
| | <i>your hand</i> | | | <i>your left hand</i> |
| (151) | roo ʔaba ʔoe | | | |
| | <i>two hand you</i> (sg) | | | |
| | <i>your two hands</i> | | | |

It clearly is not the case that in all languages the type of possessive construction is fully predictable from the semantic relation between the possessor and the possessed. On the other hand, it is equally obvious that the possessive systems of most Oceanic languages are by and large semantically based.

The view of Oceanic possessive systems adopted here is the following: In the case of the languages with indirect possessive constructions (which constitute a large majority of the Oceanic languages), we are dealing with possessive classifier systems, where one type of construction (direct) requires no classifier. Firstly, like numeral classifier systems the possessive classifier systems are semantically based. That is, they classify entities according to some semantic criteria. Secondly, like numeral classifier systems they exhibit overlap (the possibility of a noun to appear in more than one type of construction). Thirdly, like numeral classifier systems they always have a general classifier used when none of the other, more specific constructions is appropriate. Finally, like numeral classifier systems they exhibit exceptions in the types of construction in which a noun can or must appear. The only difference between numeral and possessive classifier systems lies in the nature of the classifying criteria: some properties of the entities themselves in the case of the former, and the nature of the relation between two entities in the case of the latter.

In the languages with more than one type of possessive construction but without possessive classifiers the type of construction is determined by whether the possession is alienable or inalienable. Again, there may be language-idiosyncratic exceptions and some nouns must be specified as requiring a certain possessive type under certain circumstances.

7.2 Verbal origin of the Oceanic possessive constructions?

Lynch (1982) has presented a picture of the POC possessive system radically different from that of Pawley (1973) and the one arrived at here. Lynch suggests that the present-day nominal possessive constructions derive historically from verbal structures.⁹

According to Lynch, the possessor in a present-day possessive construction derives from the object of a transitive verbal construction in POC. The possessed derives from several different kinds of element depending on the possessive type.

With kinship terms, which are reciprocal (if I am somebody's son, he is my father), the possessed derives from a verb:

father-my < *(he) *father-trans-me* (me).

('trans' is one of the transitive suffixes reconstructible for POC.)

With non-reciprocal inalienables, the possessed is the subject, and there is a 'prepositional verb' *(q)i:

hand-my < *hand (q)i-me me.

For general possession, Lynch postulates the following origin: The possessor was an indirect object, and the possessed was a direct object. There was a zero (!) ditransitive verb that carried the transitive suffix and an object suffix. The object suffix indexed the indirect object (the possessor). The subject was unspecified.

GEN -my *house* < *() \emptyset -trans-me me *house*.
CLASS Su IO DO

The food and drink possessive constructions are derived from basically the same structures as the general construction except that the original structures contain the verbs 'eat' and 'drink' respectively, both of which are ditransitive:

FOOD -my *taro* < *() *eat-trans-me* me *taro*
CLASS Su IO DO

DRINK-my *water* *() *drink-trans-me* me *water*.
CLASS Su IO DO

To derive the modern systems from their presumed POC sources, Lynch postulates a number of syntactic changes, most of which are highly unusual, to say the least. No independent evidence is offered for the assumed changes.

As far as I can tell, Lynch has three main arguments for his hypothesis:

- (i) In many languages, certain classifiers show some formal and some semantic similarities to certain verbs and therefore are cognate with these verbs. This is especially true of the alimentary/food and the drink classifiers.
- (ii) The existence of overlap between different possessive types. According to Lynch, overlap can be accounted for by assuming that different verbs could occur with the same pair of nominals.
- (iii) In some languages, all or at least some of the object and the possessive suffixes are formally identical or highly similar. I will briefly comment on the three arguments.

- (i) True, but I have argued here and elsewhere (Lichtenberk 1983a) that the so-called possessive markers are in fact classifiers. It is a well-known fact that classifiers typically derive historically from lexical morphemes. From the fact that a classifier is ultimately traceable to a verb in the proto-language, it does not follow that the nominal construction in which the classifier appears derives historically from a verbal construction in the proto-language. It is quite likely that the classifiers derive historically from (verbal) nouns.
- (ii) One need not postulate original verbal constructions to account for overlap. Overlap is not at all unusual in classifier systems. Entities may possess more than one classifying characteristic, and they can be classified on the basis of different criteria depending on the circumstances.
- (iii) True, but again an alternative and more plausible explanation is available. One common type of syntactic change is the development of independent grammatical morphemes into affixes (Givón 1971). Let us consider the pronominal forms

reconstructed for Proto-Austronesian by Blust (1977). Blust reconstructed two sets: nominative (independent) and genitive:

	nominative	genitive
1sg	i-aku	i-ku/ni-ku
2sg	i-Su (i)kaSu (polite)	i-Su/ni-Su (later on also i-mu/ni-mu as polite forms)
3sg	si-ia	i-a/ni-a
1pl excl	i-kami	i-mi/ni-mi
1pl incl	i-kita	i-ta/ni-ta
2pl	i-kamu (later also iSu)	i-mu/ni-mu (later also iSu/niSu)
3pl	si-iDa	i-Da/ni-Da

The *i* and *si* forms in the nominative set are personal articles. The *i* and *ni* forms in the genitive set are genitive prepositions (Section 6.3). It is obvious that the genitive forms are reductions of the nominative forms (a point made by Blust). It can also be assumed that the original nominative/independent forms have developed into object affixes in a number of Oceanic languages. The similarity between the possessive and the object suffixes in Oceanic languages is then due not to the former being derived from the latter but to both being derived from the same original set of independent pronouns.

A different sort of problem with Lynch's hypothesis is the fact that possessive classifiers are found in some languages that are not members of the Oceanic subgroup (see the next section). This means that unless one assumes independent parallel developments in the Oceanic subgroup and the other languages, the development from verbal to nominal constructions postulated by Lynch would have taken place at a pre-Oceanic stage.

I conclude that there is no evidence which forces us to assume a verbal origin of the present-day Oceanic possessive systems and a great deal which supports an alternative reconstruction along the lines developed here.

7.3 Are the multiple possessive construction types an innovation of Proto-Oceanic?

The question asked here is whether the multiple possessive type system as reconstructed here is an innovation of POC or whether similar systems are found elsewhere in the Austronesian family and are thus reconstructible at a higher level. The same question has been considered by Pawley (1973), who suggests that the possessive classifiers **na* and **ka* and the semantic categories they mark (general, alimentary/food and subordinate possession) are an innovation of POC.

There is evidence that systems of multiple possessive construction types, some of them with possessive classifiers, exist outside of the Oceanic subgroup.

Numfoor of West Irian has a distinction between inalienable and alienable possession. In inalienable possession, the possessed carries a possessive suffix. In alienable possession, special possessive forms (possessive pronouns?) are used.

(152) kma -mi
father-your (sg)
your father

(153) awa kobe-na
mango we (incl)
our mangoes (the suffix -na is
used when the possessed is in-
animate and plural)

Buli of South Halmahera makes a distinction between direct and indirect possessive constructions of the same sort as is found in many Oceanic languages. In indirect constructions, two possessive classifiers are used: an alimentary one *na* ~ *ra* and a general one *ni* ~ *ri* (the *ra* and *ri* alternants are used before possessive suffixes that begin with *r*).

- | | |
|---|--|
| (154) ja [?] ulo -k
I -heart-my
my heart | (155) ja [?] na -k pingē
I -CLASS-my rice
my rice |
| (156) ja [?] ni -k ebai
I -CLASS-my house
my house | (157) Masofo hnjē -∅ mtā-∅
M. mother-his eye-her
Masofo's mother's eye |
| (158) Tjangkura-kura na -∅ tela
T. CLASS-his banana
Tjangkura-kura's bananas | |
| (159) Nur ire Patendibor ri -ri fola
N. and P. CLASS-their house
Nur and Patendibor's house | |

Other languages of South Halmahera exhibit the same direct-indirect possessive construction distinction as well as the alimentary-general distinction within the indirect possessive type (Adriani and Kruijt 1914).

A distinction between inalienable and alienable possession also exists in Kaitetu of Central Maluku. In inalienable possession, possessive suffixes are added onto the possessed. In alienable possession, the corresponding elements follow independent pronouns—as suffixes in the singular, and apparently as free forms in the plural. I have no data on possessive constructions where the possessor is a noun phrase with a nominal head.

- | | |
|--|--|
| (160) ale mata-m
you(sg) eye-your
your eye | (161) ale -m luma
you(sg)-your house
your house |
| (162) ite mata-ka
we eye -our
(incl)
our eyes | (163) ite ka luma
we our house
(incl)
our house |

Similar possessive systems are found in other languages of Central Maluku (Collins 1983).

Blust (1982) suggests the subgrouping of Austronesian given in Figure 2.

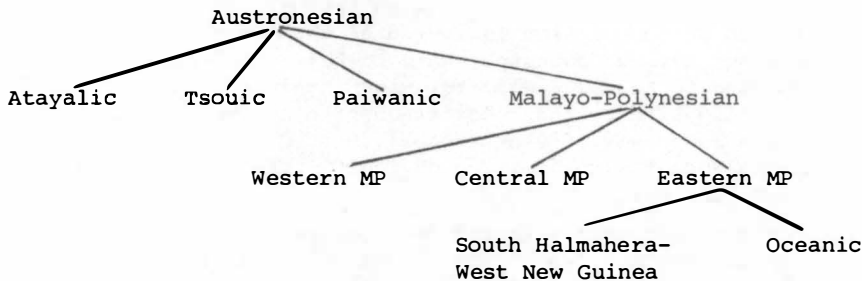


Figure 2: Subgrouping of Austronesian (Blust 1982)

Buli and the other languages that exhibit a direct-indirect possessive construction distinction belong in the South Halmahera-West New Guinea branch of Eastern MP. The languages that exhibit an alienable-inalienable possession contrast are members of Central MP (the Central Maluku languages) or the South Halmahera-West New Guinea branch of Eastern MP (Numfoor). Blust suggests that Central MP and Eastern MP may form a subgroup co-ordinate with Western MP (in fact, in another paper (1978b) Blust postulates a Central-Eastern MP subgroup).

What can we conclude from the data available? Firstly, it appears that a basic alienable-inalienable distinction must be postulated for the proto-language common to Central MP and Eastern MP. This alienable-inalienable distinction may be taken as evidence supporting Blust's Central-Eastern MP subgroup. Secondly, a direct-indirect possessive construction distinction and an alimentary-general distinction within the indirect subtype must be postulated for Eastern MP. These two types of distinction support Blust's (1978b) phonological, lexical and semantic evidence for his Eastern MP subgroup. The only innovations of POC then are the development of the drink indirect possessive type, the form of the alimentary/subordinate classifier (see below), and perhaps the category of subordinate possession.

Given these broad outlines of the pre-Oceanic stages, a new kind of problem emerges: How does one reconcile the forms of the corresponding classifiers in the South Halmahera languages and in POC? Consider the following forms:

	South Halmahera	Proto-Oceanic
alimentary	na	*ka
general	ni	*na

The *ka form is clearly an innovation of POC. The POC general classifier is formally identical with the alimentary classifier of South Halmahera. Assuming that the two forms are cognate, the question arises whether it is the function of the South Halmahera reflex that is innovative or that of the POC reflex. An answer to this question must await more data and further research.

NOTES

1. This is a slightly revised version of the paper delivered at the 15th Pacific Science Congress under the title "The syntax of possessive constructions in Proto-Oceanic". I am grateful to the following people for comments on the version delivered at the Congress: Joel Bradshaw, Ann Chowning, Raymond Johnston, Malcolm Ross and Lawrence Reid. Above all, I am indebted to Andrew Pawley for many invaluable comments on an earlier draft. I also wish to thank Geoffrey White for making available to me his data on the languages of Santa Isabel, Ross Clark for his data on Sesake, and Malcolm Ross for information on Tabar. The To'aba'ita data come from Lawrence Fo'ana'ota.
2. A list of the languages referred to, together with the subgroups they belong in and the sources of the data is given in the Appendix. The transcription conventions used here are those of the sources.

3. Geraghty (1983) makes a distinction between 'direct' and 'indirect' possessive affixes. The former are attached to the possessed, the latter are attached to 'possessive markers', i.e. classifiers in the terminology of this paper. Here the terms 'direct' and 'indirect' refer to types of possessive construction.
4. The h represents lengthening of the preceding vowel. The underlying representation of nima-h is nima -a
CLASS-*his*.
5. The Sonsorol-Tobi forms are dialectal variants.
6. Similar relationships can also be discerned in Dyen's PAN set although there they are in some cases masked by other, incompatible reconstructions.
7. I am grateful to Andrew Pawley for this suggestion.
8. Reid (1983) accepts the existence of the two *ka possessive classifiers in POC postulated by Pawley. However, he argues that the subordinate classifier had developed from a Proto-Malayo-Polynesian preposition *ka, to which oblique pronouns were attached.
9. In an earlier paper (1973), Lynch postulated a *synchronic* relationship between nominal and verbal possessive structures. He suggested that nominal possessive constructions were derived from underlying verbal structures.

APPENDIX

Languages referred to in the text, the subgroups they belong in, and the sources of the data.

PART I: OCEANIC LANGUAGES

Language	Primary subgroup of Oceanic (Figure 1, Section 6.1)	Source(s) of data
Amari	Adzera	Holzknecht 1980
Are (Mukawa)	SE Papuan	Paisawa et al. 1976
Aroma	SE Papuan	Lynch 1973
Arosi	SE Solomonian	Capell 1971
Atchin	Remote Oceanic	Capell and Layard 1980
A'ara (Maringe)	Central Isabel	White, n.d.
Babatana	Choiseul	Money, 1950s
Balawaia	SE Papuan	Kolia 1975
Banoni	Bougainville-Buka	Lincoln 1976
Bareke	New Georgia	Grace 1955
Bonkovia	Remote Oceanic	Blust 1978a
Bugotu	Central Isabel	Codrington 1885 Ivens 1933-1935a
Dehu	New Caledonia-Loyalty Is.	Tryon 1967
Dobu	SE Papuan	Capell 1943, Lithgow 1975
Duke of York Is.	Tolai-New Ireland	Codrington 1885
Efate	Remote Oceanic	Codrington 1885
Eromanga	S New Hebrides	Ray 1926
Fagani	SE Solomonian	Codrington 1885
Gedaged	Sepik-Madang	Dempwolff, n.d.
Gilbertese	Remote Oceanic?, Nuclear Micronesian?	Cowell 1950
Gitua	Rai	Lincoln, n.d.
Gog	Remote Oceanic	Codrington 1885
Hawaiian	Remote Oceanic	Wilson 1982
Houailou	New Caledonia-Loyalty Is.	Lichtenberk 1978
Iaai	New Caledonia-Loyalty Is.	Ozanne-Rivierre 1976
Iamalele	SE Papuan	Beaumont and Beaumont 1975
Iduna	SE Papuan	Huckett 1974
Kairiru	Sepik-Madang	Wivell 1981
Kaliai	Rai	Counts 1969
Kaniet	Admiralties-Western Is.	Blust 1978a
Kia	Kia	White, n.d.
Kiai	Remote Oceanic	Ludvigson, n.d.
Kosraean	Remote Oceanic?, Nuclear Micronesian?	Lee 1975
Kubokota	New Georgia	Grace 1955
Kwaio	SE Solomonian	Keesing 1985
Kwara'ae	SE Solomonian	Deck 1934
Laghu	Remote Oceanic	White, n.d.
Lakon	Remote Oceanic	Codrington 1885
Lenakel	S New Hebrides	Lynch 1978

Language	Primary subgroup of Oceanic	Source(s) of data
Leon	Remote Oceanic	Codrington 1885
Loniu	Admiralties-Western Is.	Blust 1978a
Luqa	New Georgia	Grace 1955
Maewo	Remote Oceanic	Codrington 1885
Malo	Santa Cruz	Wurm 1972
Manam	Sepik-Madang	Lichtenberk 1983b
Marovo	New Georgia	Grace 1955
Marshallese	Remote Oceanic?, Nuclear Micronesian?	Zewen 1977
Mokilese	Remote Oceanic?, Nuclear Micronesian?	Harrison 1976
Mono-Alu	Bougainville-Buka	Fagan, forthcoming
Mota	Remote Oceanic	Codrington 1885
Motu	SE Papuan	Lister-Turner and Clark n.d.
Nada	SE Papuan	Capell 1943
Nakanai	Rai	Johnston 1980
Nemboi	Santa Cruz	Wurm 1972
Nemi	New Caledonia-Loyalty Is.	Ozanne-Rivierre 1979
Nenemas	New Caledonia-Loyalty Is.	Haudricourt 1963
Nguna	Remote Oceanic	Schütz 1969
Nissan (Nehan)	Nehan (Nissan)	Todd 1978
Nooli	Santa Cruz	Wurm 1972
Paama	Remote Oceanic	Crowley 1982
Raga	Remote Oceanic	Walsh 1966
Reefs	Santa Cruz	Wurm 1972
Roviana	New Georgia	Waterhouse 1949
Sesake	Remote Oceanic	Codrington 1885, R. Clark, p.c. Capell 1969
Sonsorol-Tobi	Remote Oceanic?, Nuclear Micronesian?	Capell 1969
Standard Fijian	Remote Oceanic	Churchward 1941, Milner 1972, Pawley 1973 Capell 1943, Lynch 1973
Suau	SE Papuan	Capell 1943, Lynch 1973
Tabar	Tolai-New Ireland	M. Ross, p.c.
Tami	Coastal Huon Gulf	Bamler 1900
Tangoa	Remote Oceanic	Ray 1926
Tigak	Tolai-New Ireland	Beaumont 1979
Tolai	Tolai-New Ireland	Waterhouse 1939, Mosel 1980
Tomoip	Tumuip (Tomoip)	Johnston 1983
To'aba'ita	SE Solomonic	Lichtenberk, n.d.
Ulawa	SE Solomonic	Ivens 1913-1914
Ulithian	Remote Oceanic?, Nuclear Micronesian?	Sohn and Bender 1973
Utupua	Santa Cruz	Pawley 1973
Vanikolo	Santa Cruz	Pawley 1973
Vaturanga	SE Solomonic	Ivens 1933-1935b
Volow	Remote Oceanic	Codrington 1885
Wayan	Remote Oceanic	Pawley 1982b
Western Fijian communalects	Remote Oceanic	Geraghty 1983
Wogeo	Sepik-Madang	Blust 1978a
Yabem	Coastal Huon Gulf	Dempwolff 1939

PART II: NON-OCEANIC LANGUAGES

Language	Subgroup of Austronesian (Figure 2, Section 7.3)	Source of data
Buli	South Halmahera-West New Guinea	Maan 1951
Kaitetu	Central Malayo-Polynesian	Collins 1983
Numfoor	South Halmahera-West New Guinea	Patz 1978

BIBLIOGRAPHY

ADRIANI, N. and A.C. KRUIJT

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

BAMLER, G.

- 1900 Bemerkungen zur Grammatik der Tamisprache. *Zeitschrift für Afrikanische und Oceanische Sprachen* 5:198-216.

BEAUMONT, Clive H.

- 1979 *The Tigak language of New Ireland*. *PL*, B-58.

BEAUMONT, John and Margaret BEAUMONT

- 1975 Iamalele clause types and structures. In Loving, ed. 1975:81-152.

BLUST, Robert A.

- 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics, University of Hawaii* 9/2:1-15.
- 1978a *The Proto-Oceanic palatals*. Memoir No.43. Wellington: The Polynesian Society.
- 1978b Eastern Malayo-Polynesian: a subgrouping argument. In Wurm and Carrington, eds 1978:181-234.
- 1982 The linguistic value of the Wallace Line. *Bijdragen tot de Taal-, Land- en Volkenkunde* 138:231-250.

BRADSHAW, Joel

- 1982 Word order change in Papua New Guinea Austronesian languages. Ph.D. dissertation, University of Hawaii.

CAPELL, A.

- 1943 *The linguistic position of south-eastern Papua*. Sydney: Australasian Medical Publishing Company.
- 1969 *Grammar and vocabulary of the language of Sonsorol-Tobi*. Oceania Linguistic Monograph No.12. Sydney: University of Sydney.

- CAPELL, A.
1971 *Arosi grammar*. PL, B-20.
- CAPELL, A. and J. LAYARD
1980 *Materials in Atchin, Malekula: grammar, vocabulary and texts*. PL, D-20.
- CHURCHWARD, C. Maxwell
1941 *A new Fijian grammar*. Suva: Government Press.
- CODRINGTON, Robert
1885 *The Melanesian languages*. 2nd edn 1974. Amsterdam: Philo Press.
- COLLINS, James T.
1983 *The historical relationships of the languages of Central Maluku, Indonesia*. PL, D-47.
- COUNTS, David R.
1969 *A grammar of Kaliai-Kove*. *Oceanic Linguistics* Special Publication No.6. Honolulu: University of Hawaii Press.
- COWELL, Reid
1950 *The structure of Gilbertese*. Beru, Gilbert Islands: Rongorongo Press.
- CROWLEY, Terry
1982 *The Paamese language of Vanuatu*. PL, B-87.
- DECK, Norman C.
1934 *Grammar of the language spoken by the Kwara'ae people of Mala, British Solomon Islands*. Reprint No.5. Wellington: The Polynesian Society.
- DEMPWOLFF, Otto
1939 *Grammatik der Jabêm-Sprache auf Neuguinea*. *Abhandlungen aus dem Gebiet des Auslandskunde* 50. Hamburg: Friederichsen, De Gruyter.
n.d. *Grammar of the Graged language*. Karkar Island: Lutheran Mission.
- DYEN, Isidore
1974 *The Proto-Austronesian enclitic genitive pronouns*. *Oceanic Linguistics* 13:17-31.
- FAGAN, Joel L.
forth- *A grammatical analysis of Mono-Alu, Bougainville Straits, Solomon coming Islands*. PL, series B.
- GERAGHTY, Paul A.
1983 *The history of the Fijian languages*. *Oceanic Linguistics* Special Publication 19. Honolulu: University of Hawaii Press
- GIVÓN, Talmy
1971 *Historical syntax and synchronic morphology: an archaeologist's field trip*. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society*, 394-415. Chicago: Chicago Linguistic Society.

GRACE, George W.

- 1955 Linguistic questionnaire for Solomon Islands languages. Tri-Institutional Pacific Program.

HARRISON, Sheldon P.

- 1976 *Mokilese reference grammar*. Honolulu: University Press of Hawaii.

HAUDRICOURT, André G.

- 1963 *La langue des Nénémas et des Nigoumak (dialectes de Poum et de Koumac, Nouvelle-Calédonie)*. Auckland: Linguistic Society of New Zealand.

HOLZKNECHT, Susanne C.

- 1980 A morphology and grammar of the Amari language, Morobe Province, Papua New Guinea. B.A. Honours thesis, University of Queensland.

HOOPER, Robin

- 1985 Proto-Oceanic *qi. In this volume, 141-198.

HUCKETT, Joyce

- 1974 Notes on Iduna grammar. In Alan Healey, ed. *Three studies in languages of Eastern Papua. Workpapers in Papua New Guinea languages*, 3:63-133. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

IVENS, W.G.

- 1913- Grammar of the language of Ulawa, Solomon Islands. *Journal of the Polynesian Society* 22:28-35, 96-103, 134-140, 219-224; 23:21-27.
1914
1933- A grammar of the language of Bugotu, Ysabel Island, Solomon Islands.
1935a *Bulletin of the School of Oriental Studies* 7:141-177.
1933- A grammar of the language of Vaturanga, Guadalcanal, British Solomon
1935b Islands. *Bulletin of the School of Oriental Studies* 7:349-357.

JOHNSTON, Raymond L.

- 1980 *Nakanai of New Britain: the grammar of an Oceanic language*. PL, B-70.
1983 The continuing quest for Proto-Southwest New Britain. Paper presented to the 15th Pacific Science Congress.

KEESING, Roger M.

- 1985 *Kwaio grammar*. PL, B-88.

KOLIA, J.A.

- 1975 A Balawaia grammar sketch and vocabulary. In T.E. Dutton, ed. *Studies in languages of central and south-east Papua*. PL, C-29: 107-226.

LEE, Kee-dong

- 1975 *Kusaiean reference grammar*. Honolulu: University Press of Hawaii.

LICHTENBERK, Frantisek

- 1978 A sketch of Houailou grammar. *Working Papers in Linguistics, University of Hawaii* 10/2:74-116.

LICHTENBERK, Frantisek

- 1979 Further evidence for Proto-Oceanic *ñ. *Oceanic Linguistics* 18:171-201.
- 1983a Relational classifiers. *Lingua* 60:147-176.
- 1983b *A grammar of Manam*. *Oceanic Linguistics* Special Publication no.18. Honolulu: University of Hawaii Press.
- n.d. Notes on To'aba'ita.

LINCOLN, Peter C.

- 1976 Describing Banoni, an Austronesian language of Southwest Bougainville. Ph.D dissertation, University of Hawaii.
- 1978 Reef-Santa Cruz as Austronesian. In Wurm and Carrington, eds 1978: 929-967.
- n.d. Basics of Gitua language. Typescript.

LISTER-TURNER, R. and J.B. CLARK

- n.d. *A grammar of the Motu language of Papua* (2nd edn, edited by Percy Chatterton). Sydney: Government Printer.

LITHGOW, David

- 1975 A grammatical analysis of a Dobu text. In Loving, ed. 1975:25-56.

LOVING, Richard, ed.

- 1975 *Papers in five Austronesian languages*. *Workpapers in Papua New Guinea Languages* 12. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

LUDVIGSON, Tomas

- n.d. Silent exploitation: an exercise in "comparative ethno-economics". Typescript.

LYNCH, John

- 1973 Verbal aspects of possession in Melanesian languages. *Oceanic Linguistics* 12:69-102.
- 1978 *A grammar of Lenakel*. PL, B-55.
- 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*, vol.1: *Currents in Oceanic*, 243-268. PL, C-74.

LYONS, John

- 1968 *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

MAAN, G.

- 1951 *Proeve van een Bulische Spraakkunst*. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde, vol.10. The Hague: Martinus Nijhoff.

MILKE, Wilhelm

- 1965 Comparative notes on the Austronesian languages of New Guinea. *Lingua* 14:330-348.

MILNER, G.B.

- 1972 *Fijian grammar*. 3rd edn. Suva: Government Press.

MONEY, Lucy H.

- 1950s Babatana language. Typescript.

MOSEL, Ulrike

- 1980 *Tolai and Tok Pisin: the influence of the substratum on the development of New Guinea Pidgin*. *PL*, B-73.

OZANNE-RIVIERRE, Françoise

- 1976 *Le Iaai, langue mélanésienne d'Ouvéa (Nouvelle-Calédonie)*. Langues et civilisations à tradition orale, 20. Paris: Société d'études linguistiques et anthropologiques de France.

- 1979 *Textes Nemi (Nouvelle-Calédonie)*, vol.1: *Kavatch et Tendo*. Langues et civilisations à tradition orale, 31. Paris: Société d'études linguistiques et anthropologiques de France.

PAISAWA, Elsie, Louise PAGOTTO and Joan KALE

- 1976 *A short sketch of Are (or Mukawa), Milne Bay, Papua New Guinea*. Occasional Paper No.2. Department of Language, University of Papua New Guinea.

PATZ, Elisabeth

- 1978 The case marking and role coding system of Numfoor-Biak. *Oceanic Linguistics* 17:141-161.

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 P. Bishop Museum.

- 1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103-188.

- 1977 Redefining Eastern Oceanic. Typescript.

- 1978 The New Guinea Oceanic hypothesis. *Kivung* 11:99-151.

- 1982a Rubbish-man commoner, big-man chief? Linguistic evidence for hereditary chieftainship in Proto-Oceanic society. In Jukka Siikala, ed. *Oceanic studies; Essays in honour of Aarne A. Koskinen*, 33-52. Transactions of the Finnish Anthropological Society No.11. Helsinki: Finnish Anthropological Society.

- 1982b Notes on the marking of possessive relations in Wayan. Typescript.

RAY, Sidney H.

- 1919 The Melanesian possessives and a study in method. *American Anthropologist* 21:347-360.

RAY, Sidney H.

- 1926 *A comparative study of the Melanesian Island languages.* Cambridge: University Press.

REID, Lawrence A.

- 1983 The sources of POC genitive constructions. Paper presented at the 15th Pacific Science Congress.

SCHÜTZ, Albert J.

- 1969 *Nguna grammar.* *Oceanic Linguistics* Special Publication No.5. Honolulu: University of Hawaii Press.

SCHÜTZ, Albert J. and Tevita NAWADRA

- 1972 A refutation of the notion "passive" in Fijian. *Oceanic Linguistics* 11:88-109.

SIMONS, Gary F.

- 1980 Morphological reconstruction and change in the pronoun systems of the Malaitan languages. Paper presented at the 3rd New Zealand Linguistics Conference, September 3-5, Auckland.

SOHN, Ho-min and B.W. BENDER

- 1973 *A Ulithian grammar.* PL, C-27.

TODD, Evelyn M.

- 1978 A sketch of Nissan (Nehan) grammar. In Wurm and Carrington, eds 1978:1181-1239.

TRYON, D.T.

- 1967 *Dehu grammar.* PL, B-7.

WALSH, David S.

- 1966 The phonology and phrase structure of Raxa. M.A. thesis, University of Auckland.

WATERHOUSE, J.H.L.

- 1939 *A New Guinea language book (Blanche Bay dialect).* Sydney: Australasian Medical Publishing Company.
 1949 *A Roviana and English dictionary* (revised and enlarged by L.M. Jones). Sydney: Epworth.

WHITE, Geoffrey M.

- n.d. Notes on Santa Isabel languages. MS.

WILSON, William H.

- 1976a The O and A possessive markers in Hawaiian. M.A. thesis, University of Hawaii.
 1976b The o/a distinction in Hawaiian possessives. *Oceanic Linguistics* 15:39-50.
 1982 *Proto-Polynesian possessive marking.* PL, B-85.

WIVELL, Richard

1981 Kairiru grammar. M.A. thesis, University of Auckland.

WURM, S.A.

1972 Notes on the indication of possession with nouns in Reef and Santa Cruz Islands languages. In *Papers in linguistics of Melanesia*, No.3:85-113. PL, A-35.

1978 Reefs-Santa Cruz: Austronesian, but ...! In Wurm and Carrington, eds 1978:969-1010.

WURM, S.A. and Lois CARRINGTON, eds

1978 *Second International Conference on Austronesian Linguistics: proceedings*, fascicle 1: *Western Austronesian*; fascicle 2: *Eastern Austronesian*. PL, C-61.

ZEWEN, François-Xavier N.

1977 *The Marshallese language: a study of its phonology, morphology, and syntax*. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg, vol.10. Berlin: Dietrich Reimer.

PROTO-OCEANIC *qi

Robin Hooper

1. INTRODUCTION

In a recent paper, "Proto-Austronesian genitive determiners", Lawrence Reid (1981:98-104) reconstructs a Proto-Austronesian (PAN) genitive phrase of the form Noun $-\{n_i\}$ - Noun. Adding to evidence presented by Blust (1974, 1977) he suggests that *i and *ni were phonologically conditioned alternants, *ni occurring after vowel-final nouns, and *i after consonant-final nouns.

The reconstruction of PAN *ni has been accepted by linguists since Dempwolff on the basis of widespread reflexes, but less attention has been paid to *i. The evidence assembled by Reid and Blust in their papers is taken largely from Western Austronesian languages. There is in fact far more widespread evidence for a genitive linking particle *i (or *qi) in Oceanic languages than is indicated by the examples from Oceanic quoted by Blust (1977:4). However Oceanic witnesses provide no indication whatsoever that this linker was merely a phonological variant of *ni.

Pawley (1972, 1973) reconstructed a particle *(q)i for Proto-Eastern Oceanic and more tentatively for Proto-Oceanic (POC), and suggested that it was used to mark inalienable possession. The evidence he cited was of three kinds:

1. The use of i before suffixed pronominal possessors in Standard Fijian and in Kuanua (1973:158), and before personal names (1972:34; no languages cited).
2. The suffix -i or -gi on 'independent' forms of nouns that are usually inalienably possessed in Northern New Hebrides-Banks Islands languages, e.g. Mota mata-i *eye* (1972:115; 1973:158).
3. A linking particle in phrases embodying certain numeral classifiers such as *mata or *pua, indicating an item in a group or series. This particle has the form (')i or (')e in the South-East Solomon Islands, and 'i in Tongan, pointing to POC classifiers *mataqi and *puaqi (1972:34,109).

Now these are three rather different structures. The first and second share the notion of inalienable possession; the first and third share the use of i or 'i to link two nouns, though in very different kinds of phrase. More evidence is required before one can reach any firm conclusion on the distribution and function of *qi in the genitive system of POC, and the nature of the contrast, if any, between *qi and *ni. The object of this paper is to supply that evidence by examining possible reflexes of *qi in a number of Oceanic languages.¹

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 141-167. *Pacific Linguistics*, C-88, 1985.

© Robin Hooper

After some notes on phonology in section 2, I discuss in section 3 the use of *i* as a genitive preposition with personal noun and pronoun possessors in a number of languages. Section 4 deals with the classifiers found in Tongan and South-East Solomons languages, and the widely dispersed idiomatic expressions derived from these classifiers. Next, in section 5, I examine the function of reflexes of **qi* in the genitive systems of certain Solomon Islands and Vanuatu languages, and finally in section 6, examples in which **qi* is reflected in certain idiomatic or frozen collocations, suggestive of earlier more widespread use, notably the 'independent suffix' of Banks Islands languages. After stating my conclusions in section 7, I briefly discuss the relationship between **qi* and **ni*, and the possibility that the **qi* - **ni* contrast existed in PAN.

Throughout this paper I use the terms alienable and inalienable possession in the familiar way, but with the understanding that the domains referred to by these terms will differ slightly from language to language.

All examples are given in the orthography of the sources, except in the case of Lonwolwol, where I have written Paton's *ε* as *e* and *ο* as *o*.

The following list gives abbreviations of language names cited in the text, and the main sources for those languages:

ARS	Arosi	Capell 1971
BUG	Bugotu	Ivens 1933
FIJ	Standard Fijian	Churchward 1941, Capell 1941
KWM	Kwamera	Ray 1926
KWO	Kwaio	Keesing 1975
KWR	Kwara'ae	Deck 1933
LIF	Lifu	Ray 1926
LON	Lonwolwol	Paton 1971
LUA	Luangiua	Salmond 1976
MTA	Mota	Codrington 1885
NUK	Nukuoro	Carroll 1973
OBA	Aoba	Ivens 1940a, Codrington 1885
PAA	Paama	Crowley 1982
REN	Rennell	Elbert 1975
SAM	Samoan	Pratt 1911, Milner 1966
TIG	Tigak	Beaumont 1979
TOL	Tolai (Kuanua)	Franklin 1962, Mosel 1977
TON	Tongan	Churchward 1953, 1959
TSR	Tasiriki	Ray 1926

2. PHONOLOGY

I shall assume the correctness of the reconstruction **qi* as the POC form of a linking particle cognate with at least some of the forms cited by Pawley (1972 and 1973). There are however some irregularities not accountable for in terms of the established sound correspondences.

2.1 Variation in the consonant

The chart below shows reflexes of POC **q* and **k* in a number of Oceanic languages:

POC	MTA	MRL	KWR	PPN	TON	SAM
*q	∅	∅	∅	?	?	∅
*k	g	g	k	k	k	?

(g = velar fricative)

POC *q disappears from all daughter languages except Proto-Polynesian, where it is reflected as glottal stop. This phoneme is retained in Tongan and a small number of other Polynesian languages. The form 'i which occurs in Tongan is thus a regular reflex of *qi. However this form 'i also appears in some languages in which *q is regularly lost, for example in Kwaio and Kwara'ae, and in Kwara'ae genitive 'i is phonemically distinct from the locative preposition i and the feminine personal article i. Some Samoan collocations involving the classifiers also unexpectedly exhibit glottal stop, for example fua'i fai a *single banana*.

North Vanuatu languages show variation in the form of the independent noun suffix between Mota -i, Merlav -i ~ -gi, and Lakon -gi ~ -n. (The -n suffix probably derives from a different particle, and will not be discussed here.) The conditions governing these alternations are not clear from Codrington's data, but g is a regular reflex of *k and a plausible if irregular reflex of *q.

An alternative hypothesis, that POC had a genitive particle with variants *qi, *ki, would account for the occurrence of -gi in North Vanuatu languages, and 'i in Samoan. However the rarity of a reflex ki in Oceanic languages makes this an unsatisfactory solution. Schütz (1969:41) describes Nguna genitive prepositions ki and ni, with functions similar to o and a; and in the Arosi counting system the following forms occur: rau ki haru *ten thousand* (lit. *leaf-ki-tree*), rawa i niu *one hundred thousand* (lit. *leaf-i-coconut*), and raurau ni ha'aro *one million* (lit. *leaf-ni-tree*) (Capell 1971:52).

In view of the erratic nature of this variation in the consonant phoneme, it seems wisest to assume a sporadic change of the form *q > k.

2.2 Variation in the vowel

The forms 'e and e occur in several South-East Solomon Islands languages, not as the only reflex of *qi, but alternating with 'i or i. In Kwaio the phonological conditions are clear: 'i occurs after nouns with final high vowels, 'e after mid and low vowels. There are exceptions, such as fo'i. (See Keesing 1975: xxx-xxxii.) In other languages the variation on present evidence seems unpredictable, as in the Kwara'ae classifiers ma'e, fa'i and gwa'i (Deck 1933: 8-9).

In Standard Fijian, by a fortunate accident of phonological history, POC *(q)i *locative preposition* (Pawley 1972:85) is reflected irregularly, but consistently, as e, and is thus clearly differentiated from surviving examples of POC genitive *qi, which is reflected as i:

- (1) e mua i liu
 loc. *front* of *forward*
 at the bow (of a ship)

In Tongan, Samoan, and some Polynesian Outlier languages, some collocations involving reflexes of the classifier *mataqi have the form mataa- instead of the expected mata'i- or matai-. Reasons for this variation are discussed below, section 4.1.

3. PERSONAL NOUN POSSESSORS WITH i

In a few languages i occurs in noun-genitive-noun phrases before what I shall call personal noun possessors (PNP). The set of nouns designated by this term varies from language to language, but always includes proper nouns which are names of persons. The languages considered here are Lifu from Loyalty Islands, Standard Fijian, and two Tolai-New Ireland languages, Tigak and Tolai. Table I shows the structure of genitive phrases with personal noun possessors and common noun possessors (CNP) in these languages.

3.1 Lifu

In this language, PNP refers to proper nouns and to words like 'chief' and 'father' when used to refer to individuals (Ray 1926:119). In Lifu, i occurs in all genitive phrases in which the possessor is a PNP, and also before second and third person pronoun possessors. (First person pronoun possessors are indicated by a suffix on the head noun.) No distinction appears to be made between alienable and inalienable possession:

- | | |
|---|---|
| (2) la he i loane
art. head of I.
<i>John's head</i> | (3) la nekö i joxu
art. son of chief
<i>the chief's son</i> |
| (4) la uma i kaka
art. house of father
<i>father's house</i> | (5) la uma i 'ö
art. house of
<i>your house</i> |
| (6) la ite inyö i angate
art. pl. teeth of
<i>their teeth</i> | |

In the case of CNPs, part nouns and other inalienable nouns are followed directly by the genitive noun, as in (7) and (8), whereas alienable nouns are followed by the genitive preposition ne, as in (9) and (10):

- | | |
|--|---|
| (7) la hni ate
art. heart man
<i>the heart of man</i> | (8) la wene la nöjei sinöe
art. fruit art. pl. tree
<i>the fruit of the trees</i> |
| (9) la uma ne la ate
art. house of art. man
<i>the man's house</i> | (10) la nöjei sinöe ne la hlapa
art. pl. tree of art. garden
<i>the trees of the garden</i> |

3.2 Standard Fijian

In Standard Fijian, PNP refers to proper names of persons. The possessive marking before PNPs reflects the complexity of the pronominal possessive system. Four distinct possessive markers occur: i (inalienable possession), nei (general

possession), *kei* (alimentary and subordinate possession), and *mei* (drinkable possession). The last three can be analysed into a possessive classifier (see Lichtenberk, in this volume) plus *i*:

- | | |
|---|--|
| (11) na ulu i Cakobau
art. head of C.
Cakobau's head | (12) na vale nei Pita
art. house class-of P.
Peter's house |
| (13) na uvi kei Pita
art. yam class-of P.
Peter's yam | (14) na sucu mei Pita
art. milk class-of P.
Peter's milk |

In addition, in the paradigm of suffixed possessive pronouns used to indicate inalienable possession, *i* appears before non-singular first person exclusive forms only (Pawley 1973:158; Churchward 1941:27):

- (15) na tina i-keimami
art. mother 1 pl.excl
our mother

CNPs, if animate, normally follow the appropriate pronominal possessive construction:

- (16) na no-dratou waqa na cauravou
art. class-their canoe art. young man
the young men's canoe

However a prepositional construction with *ni* is also available to CNPs, and this neutralises the distinction between the different types of possession as well as expressing genitive relationships other than possession:

- | | |
|---|---|
| (17) na yaca ni gone oqō
art. name of child this
the child's name | (18) na vale ni vuli
art. house of learn
the school house |
|---|---|

Note also that a non-productive pattern of genitive phrases with a linking particle *i* before CNPs does occur in Fijian. I will discuss these in a later section.

3.3 Tigak

In Tigak, the PNP category includes proper nouns and kinship nouns. If the genitive phrase expresses inalienable possession, the PNP is introduced by *i* and the head noun carries a possessive suffix:

- | | |
|---|---|
| (19) tang lingi-na i Gamsa
art. voice-his of G.
Gamsa's voice | (20) na tiga -na i Gamsa
art. brother-his of G.
Gamsa's brother |
| (21) a patu-na i tiga -na
art. head-his of brother-his
his brother's head | |

If the genitive phrase expresses alienable possession, the PNP is introduced by *i* or *te*:

- (22) ta sasapulai i tiga -k
art. payment of brother-my
payment for my brother

- (23) tang lui te Makeo
 art. *house of M.*
Makeo's house

For CNPs, the genitive noun directly follows the suffixed noun in cases of inalienable possession, as in (24), or is introduced by the prepositions *ina* or *tana* in cases of alienable possession, as in (25) and (26):

- (24) tang lingi-na tang ulina gura
 art. *voice-her art. woman this*
this woman's voice
- (25) a aisok tana vap (26) mamana ot ina masut
 art. *work of people* pl. *thing of bush*
people's work *things of the bush*

(The distinction between *i* and *ina* on the one hand and *te* and *tana* on the other, is not clear from the few examples in Beaumont 1979.)

3.4 Tolai

In Tolai (Kuanua) PNP refers to human nouns. In this language the PNP category selects *i* when inalienable possession is expressed, and *kai* in cases of alienable possession. *Ka* is the possessive morpheme or classifier to which pronominal possessive suffixes are attached, and I would analyse *kai* as composed of *ka* plus *i*. This construction is thus similar to the corresponding one in Standard Fijian:

- (27) a nuknuk i To Kadik (28) a kak i ra tutana
 art. *thought of T. K.* art. *leg of art. man*
the plan of To Kadik *the man's leg*
- (29) lima i dat (30) a pal kai ra tutana
 art. *hand of 1 pl.incl* art. *house class-of art. man*
our hands *the man's house*
 (cf. *ka-ugu pal my house*)

CNPs, on the basis of the data available to me, must be divided into animate and non-animate. Part-nouns with animate but non-human possessors are followed directly by the CNP, as in (31) and (32), whereas inanimate possessors in phrases expressing product, substance or part-whole relationships are introduced by the connective particle *na*, which I have here glossed *of* although it has a number of other functions:

- (31) a ulu ra eu (32) a kau ra pap
 art. *head art. fish* art. *leg art. dog*
the head of the fish *the dog's leg*
- (33) a mapi na davai (34) a kap na tava
 art. *leaf of tree* art. *cup of water*
the leaf of a tree *the cup of water*

A gap in this set of data for Tolai is the category of alienable possession with common noun possessors. This may be a function of the fact that PNP in Tolai includes all human nouns. Relationships which can be unequivocally classed as alienable do not occur freely with entities other than human possessors.

3.5 Conclusions

In these four languages, PNPs are preceded by *i*, whereas CNPs are not. More significantly, it is apparent that it is the genitive noun which selects *i*, not the head noun nor the type of possession. Pawley (1972:58) reconstructs a POC personal and pronominal article **i*, and a possible connection between this article and the use of *i* before personal nouns and pronouns seems an obvious possibility. However, in none of these languages does *i* occur as a form of the personal article. I will leave this question for the moment and return to it in section 6.

Table 1			
Structure of noun-genitive-noun phrases in Lifu, Standard Fijian, Tigak and Tolai.			
LIFU	PNP		CNP
	Inalienable	NP <i>i</i> NP	NP NP
	Alienable	NP <i>i</i> NP	NP <i>ne</i> NP
STANDARD FIJIAN		PNP	CNP
	Inalienable	NP <i>i</i> NP	NP <i>ni</i> NP
	Alienable	NP <i>nei</i> <i>kei</i> NP <i>mei</i>	NP <i>ni</i> NP
TIGAK		PNP	CNP
	Inalienable	NP-suff. <i>i</i> NP	NP NP
	Alienable	NP <i>i/te</i> NP	NP <i>ina/tana</i> NP
TOLAI		PNP	CNP
	Inalienable	NP <i>i</i> NP	NP NP (animate) NP <i>na</i> NP (inanimate)
	Alienable	NP <i>kai</i> NP	

4. THE CLASSIFIERS

Pawley (1972:35,59 and 109) reconstructs three numeral classifiers for Proto-Eastern Oceanic: *po-qi and *pua-qi, evidently non-contrasting variants "selected by nouns denoting spherical objects", and *mata-qi, glossed *individual unit of series or class*. Reflexes of these classifiers are found in Polynesian (PN) languages, particularly Tongan, and in Cristobal-Malaitan languages of South-East Solomonic.

4.1 Polynesian

The classifier functions referred to by Pawley are directly reflected in the following Tongan examples, showing special counting forms used for different commodities:

- (35) ha fo'i niu / 'ufi / au 'e taha
 art. class. *coconut yam piece of thatch* num.part. *one*
one coconut / yam / piece of thatch
- (36) ha mata'i ika 'e taha
 art. class. *fish* num.part. *one*
one fish
- (37) fua'i ika
 class. *fish*
fish roe
- (38) fua'i 'ufi
 class. *yam*
fruit-like excrescences of yam plant

In addition, Tongan has other classifiers which incorporate the particle 'i: taua'i *pair of*, and lau'i *sheet of*, the latter applied to flat objects (cf. lau *leaf*):

- (39) ha taua'i niu 'e hiva
 art. class. *coconut* num.part. *nine*
nine pairs of coconuts
- (40) ha lau'i papa
 art. class. *timber*
a plank
- (41) ha lau'i sio ata
 art. class. *glass*
a sheet of glass

Example (41), and (42)-(44) below, show that this structure has been productive in post-contact times:

- (42) mata'ifika
 class-*figure*
numeral
- (43) mata'itohi
 class-*write*
letter of alphabet
- (44) mata'ipeni
 class-*pen*
pen nib

Compounds of similar structure occur in Samoan, although reflexes of these classifiers are not a productive part of the counting system as in Tongan:

- (45) mata'inūmela
 class-*number*
numeral
- (46) fua'ifa'i
 class-*banana*
a single banana

- (47) fuaitau
class- ?
a line of a song

In both Tongan and Samoan, mata'i- and a variant matā- occur in a wide range of idiomatic expressions. The erratic nature of this variation is clear from the following examples:

(48)	TONGAN	SAMOAN	
	mata'italo	matātalo	<i>taro tops for planting</i>
	mata'ihuhu	matāsusu	<i>nipple</i>
	matātangata	mataitagata	<i>fine-looking man</i>
	mata'ikoloa	mata'i'oloa	<i>most highly esteemed item</i>
	matāfanga	matāfaga	<i>beach</i>

There are a number of possible explanations of the matā- forms (which are prevalent in PN Outlier languages), none of them entirely satisfactory:

1. i has been assimilated to the preceding vowel, with concomitant lengthening of the a. This does not seem a particularly natural sound change.
2. The long ā may represent assimilation of the PN genitive prepositions a or o. It is quite plausible that the PN genitive particle should be substituted for 'i or i, of which the function is opaque in PN, but not at all apparent why this should have happened in some cases but not in others.
3. The matā- forms are nominal compounds of the kind that are common throughout Oceanic, composed of head noun + modifier noun. Although lengthening of the final vowel of the first element in a compound is not a general rule in PN there are cases where this occurs. For example, in Tokelau suffixed nominalisations of transitive verbs, if compounded with an incorporated object, regularly show such vowel-lengthening:

- (49) vali *to paint*, valiga *painting*, valigāfale *housepainting*

This explanation seems the most satisfactory. See also my comment on Kwara'ae nominal compounds below, end of section 5.1.

The semantic range of Tongan and Samoan mata- compounds embraces *point* or *prominent part of something*, where the connection with mata *eye* is most apparent, as in the words for *nipple* and *taro-tops*; or *foremost and best*, as in Samoan mataitōga *the most valuable fine mat of a collection*. Other terms are semantically opaque, suggesting considerable antiquity:

- (50) SAM mata'ifale
house
incest

Cognates of many of these forms, and other very similar compounds, occur in PN Outlier languages:

(51)	NUK	LUA	TON	REN	
	madaa ua	makaa ko'i	mata'iuha		<i>raindrop</i>
	madaa moni	makaa va'a		mataabaka	<i>canoe prow</i>
	NUK hua i gadea	<i>side of canoe away from outrigger</i>			
	REN huaa ga'akau	<i>fruits</i>			

4.2 Cristobal-Malaitan

These classifiers are retained in the Cristobal-Malaitan languages, along with a number of other classifying particles which incorporate the linker (')i or its variant (')e (see section 2.2). The following examples are representative:

- (52) *mata-qi
- | | | |
|-----|-------------|--------------------------------------|
| KWR | ma'e 'ai | <i>stick</i> |
| | ma'e fera | <i>a particular space in a house</i> |
| | ma'e fata'a | <i>a word</i> |
| KWO | maa'e mae | <i>descent group</i> |
| | maa'e obi | <i>wrist length of shell money</i> |
| SAA | maai nume | <i>house door</i> |
| | maai deni | <i>a wind</i> |
- (53) *po-qi
- | | | |
|-----|------------------|---------------------------|
| KWR | fa'i nui | <i>a coconut</i> |
| | fa'i dangi | <i>a particular day</i> |
| | fita fa'i leka'a | <i>how many journeys?</i> |
| KWO | fo'i bata | <i>a small bead</i> |
| | fo'i ka'o | <i>a bamboo internode</i> |
- (54) *pua-qi
- | | | |
|-----|-----------|----------------|
| ARS | hua i i'a | <i>a fish</i> |
| | hua i hau | <i>a stone</i> |
- (55) KWR gwa'i fau *big boulder* (cf. gwa head)
gwa'i salo *cumulus cloud*
- | | | |
|-----|-------------|---|
| KWO | tofu'i boo | <i>piece of pork</i> (cf. tofu cut) |
| | fuu'i wane | <i>group of men</i> (cf. fu'u bunch) |
| | lafa'e alo | <i>portion of taro</i> (cf. lafa portion) |
| SAA | polo i ha'a | <i>strand of shell money</i> |
| | pwire i eu | <i>length of bamboo</i> |
| | mola i uhi | <i>10,000 yams</i> |
| | rau i helu | <i>10,000 coconuts</i> |
| ARS | kora i uhi | <i>Haliotis shell</i> (cf. kora round object) |
| | kora i hau | <i>skull</i> |

Whereas the particle 'i which occurs in the Tongan classifiers is not a productive element in the grammar of that language, Cristobal-Malaitan classifier phrases with (')i are examples of a structure that is part of the genitive system of a number of South-East Solomonian languages. This structure is examined in the next section.

5. GENITIVE PHRASES IN SOME SOUTH-EAST SOLOMONIC AND NORTH AND CENTRAL VANUATU LANGUAGES

In certain South-East Solomonian and North and Central Vanuatu languages, reflexes of the particle *qi occur in phrases of a very different type from those of the languages considered in section 3, and frequently contrasting

with another genitive particle, *ni*. Ivens' attempts to distinguish *i* from *ni* in languages where both occur are all confusion. Capell, writing of Arosi, refers to "the relators *i* and *ni*, *of*, between which there does not seem to be any meaning difference, *gare i noni child of man ... madora i dangi time of day*. These nouns seem to belong to the part-whole sub-group, and perhaps *ni* forms are just a little more separate ... *dangi ni mamaro day of rest* (as against the above *time of day* which is inherent in the day, while *rest* is not)" (1971:61).

The most complete and coherent published description of the genitive system of a South-East Solomonian language is that of Deck for Kwara'ae. Moreover it describes a contrast between *i* and *ni*. I start with a summary of his account (Deck 1933:11, 15-16).

5.1 Kwara'ae

In inalienable relationships such as part-whole or body part, the possessed noun carries the third person singular possessive suffix *-na*, and the genitive noun follows directly, as in (56) and (57):

- | | | | |
|------|--|------|---|
| (56) | 'ae-na sa Pita
leg-suff. art. P.
Peter's leg | (57) | rara -na 'ai
branch-suff. tree
the branch of the tree |
|------|--|------|---|

If an alienable noun has an animate possessor, the genitive noun follows the possessed noun directly, without any intervening suffix or particle. This construction expresses true possession or ownership:

- (58) luma sa Pita
house art. P.
Peter's house

When the genitive noun is inanimate, an alienable noun is followed by the particle *ana*, composed of the possessive classifier *a-* and the third person singular suffix:

- (59) buka ana isufuta'a
book class-suff. genealogy
the book of the genealogy

Fourthly, the genitive preposition *ni*, says Deck, is used when the head noun does not take the possessive suffix *-na*, that is, is classed as alienable. "This preposition *ni* ... expresses not possession, but association or designation":

- | | | | |
|------|--------------------------------------|------|--|
| (60) | wae ni rao
man of work
servant | (61) | 'aba fola ni lukata'ilana
class paper of divorce
bill of divorcement |
|------|--------------------------------------|------|--|

Finally, according to Deck, "the genitive preposition *'i* ... becomes the substitute for the possessive suffix *-na* when abstract expressions are desired":

- | | | | |
|------|---|------|---|
| (62) | 'ae 'i wae
leg of man
human leg | (63) | mafolo 'i 'ai
block of tree
block of wood |
| (64) | fakala 'i kwaro
egg of k. shell
pearl | | |

Two points of interest emerge. Firstly, *ni* is associated with alienable relationships, and *i* with inalienable relationships. Secondly, nouns introduced by *i* or *ni* are non-specific, and function purely as modifiers of the head noun. I think this is what Deck means by the phrase "abstract expressions". Compare example (56) with (62), and (57) with (63). A variety of relationships may be expressed by these genitive phrases: body parts, substance, product or attribute, but the genitive noun is always non-specific or generic, and the whole phrase is a lexicalised compound referring to a recognised entity. Nominal compounds with the same connotations also occur in Kwara'ae, and Deck provides the following comparisons:

- | | | | | |
|------|------------------|--|-----|--|
| (65) | maga 'ai | | cf. | maga-na 'ai |
| | <i>seed tree</i> | | | <i>seed-suff. tree</i> |
| | <i>seed</i> | | | <i>the seed of a tree under discussion</i> |
| (66) | ngisingisi 'ai | | cf. | ngisingisi-na 'ai |
| | <i>chip tree</i> | | | <i>chip -suff. tree</i> |
| | <i>chips</i> | | | <i>chips of a particular tree</i> |

Deck explains these nominal compounds as cases where the 'i is dropped "though originally there". It seems more likely that both structures are available to speakers, and that in different cases one or the other form has become lexicalised.

5.2 Bugotu

Bugotu genitive phrases collected from an informant also show clearly a contrast between possession by a specific possessor, indicated by the suffix *-ña* on the head noun, and non-specific possession, indicated by the use of the linker *i*. The *i* form does not however appear to be used with human possessors:

- | | | | | |
|------|------------------------------|--|-----|------------------------------|
| (67) | ulu -ña fei | | cf. | ulu i fei |
| | <i>head-suff.</i> | | | <i>head of fish</i> |
| | <i>the head of the fish</i> | | | <i>a fish head</i> |
| (68) | legu -ña hore | | cf. | legu i hore |
| | <i>stem-suff. canoe</i> | | | <i>stem of canoe</i> |
| | <i>the stem of the canoe</i> | | | <i>canoe-stem</i> |
| (69) | kaukau-ña na mane | | cf. | kaukau mane (+kaukau i mane) |
| | <i>finger-suff. man</i> | | | <i>finger man</i> |
| | <i>the man's finger</i> | | | <i>human finger</i> |

Ivens (1933:150) gives what he calls "abstract" genitive phrases from Bugotu, such as *na aheahe i* have *the breath of life*. He comments that the particle *ni* occurs only rarely in Bugotu, whereas in some other South-East Solomonic languages, such as Nggela and Longgu, it appears to have supplanted *i*. It seems probable that a clear semantic distinction which formerly obtained between ('*i*) and *ni* has become neutralised in many languages.

5.3 Lonwolwol

Relevant data for the Vanuatu languages have not been easy to find. There is evidence of a tendency for the reflex of **qi* to disappear from the environments where we might expect it. The contrast between specific and non-specific

genitives is fully documented in the case of Lonwolwol (Paton 1971:32-34). This language, like others of the region, has lost the final vowel of the third person singular possessive suffix, which is reflected as -n. Inalienable possession by specific possessors is expressed, as in the languages considered above, by the use of this suffix on the head noun:

- (70) *hela -n vanten*
brother-suff. man
the man's brother

However there are also phrases in which the head noun is followed directly by the possessor noun: "These shorter forms are found to be compact compound nouns almost technical terms, denoting things which are entities by their own right, and not to be analysed as denoting things or persons as related to other things or persons" (Paton 1971:30). Note the following comparisons:

- | | | | |
|------|-----------------------------|-----|--------------------|
| (71) | <i>neti-n vanten</i> | cf. | <i>neti vanten</i> |
| | <i>baby-suff. man</i> | | <i>baby man</i> |
| | <i>the man's son</i> | | <i>baby</i> |
| (72) | <i>alu -n barbar</i> | cf. | <i>alu barbar</i> |
| | <i>skin-suff. pig</i> | | <i>skin pig</i> |
| | <i>the pig's hide</i> | | <i>pigskin</i> |
| (73) | <i>meta-n mak susu</i> | cf. | <i>mete susu</i> |
| | <i>eye -suff. my breast</i> | | <i>eye breast</i> |
| | <i>my nipple</i> | | <i>nipple</i> |

I think we can assume that the compound forms show phonetic loss of a linking *i*. Two kinds of evidence support this view. Firstly, loss of genitive *i* appears to be a synchronic process in Aoba and Paama, as we shall see shortly. Secondly, the modification of the second vowel in *mete susu* (cf. *meta-eye*) is consistent with assimilation to a following *i*. Note the following Lonwolwol dictionary entries:

- (74) *vera- hand of* (suffix-taking)
vere- handle (compound forming)
vere woh paddle handle

I do not have sufficient data to know if this is a regular process in Lonwolwol, but it is similar to the assimilation which takes place in Kwaio genitive phrases (see section 2.2 above), of which the following is a typical example:

- (75) *KWO lafa portion lafe'e alo portion of taro*

The Lonwolwol cognate of *ni* is *ne*, and as in Kwara'ae it occurs after a head noun which is not "suffix-taking", that is, after alienable nouns, with meanings *of*, *to do with*, or *for the purpose of*:

- (76) *gehan ne tel*
work of garden
gardening work
- (77) *we ne mūnan*
water of drink
drinking water

To summarise, the genitive noun in both these types of phrase is non-specific, and the semantic contrast between the two types is the same as that between Kwara'ae *i*-phrases and *ni*-phrases.

5.4 Aoba

In their descriptions of Aoban, both Ivens and Codrington refer to the optional use of the particle *i* between the two nouns of a genitive phrase:

- (78) qatu boe, qatu i boe
 head pig
 pig's head

Ivens also gives a disorganised account of phonological modification of the final vowel of the head noun of the genitive phrase, which in every case is compatible with assimilation to a following *i* (1940:349).

Codrington makes the following very relevant comment: "If the hand of a definite man, or the wing of a definite bird, is in view the Pronoun of the Third Person is suffixed to the former Noun" (1885:422). Here again we see the suffixed form for specific possessors contrasted with non-specific or generic genitive phrases, some of which exhibit the linker *i*.

5.5 Paama

More detailed evidence of the instability of *i* in Vanuatu languages is provided by Crowley's (1982) study of Paama. Inalienable possession by animate possessors is marked by the suffix *-n* on the head noun. Inalienable relationships in which the genitive noun is inanimate, that is the part-whole, product and substance relationships with which we are now familiar, are expressed by juxtaposition of head noun and genitive noun, except in a certain restricted set of phonological environments in which there is a linking particle *i*:

- (79) mete i tan
 eye ground
 mound made over planted yam
- (80) valenge i hat
 hollow rock
 cave
- (81) soko i āv
 remains firewood
 firewood chips

Crowley devotes some space to the mystery of this *i* (1982:91,103f). He refers to the independent noun suffix of North Vanuatu languages, and to the use of *i* with proper noun possessors in Fijian, but appears unaware that *i* is an element in the genitive system of several North Vanuatu and South-East Solomonian languages, with a function similar to that of Paama *i*. We can assume that in Paama it was once used in a wider range of non-specific inalienable genitive phrases, although now retained in only a few phonological environments.

5.6 Miscellaneous

Finally, fragments of relevant information are to be found in Ray (1926). Ray's analysis of genitive phrases in Tasiriki is curious: in the expression of inalienable possession, *ni* is the particle used if the *genitive noun* is also inalienable, *i* if the genitive noun is alienable. This dubious rule is not in

fact supported by his examples, which do however suggest a contrast between specific and non-specific possessors, the former marked with *ni*, the latter with *i* (1926:374):

- (82) 'ere ni tama -mim
face of father-2 sg.
your father's face
- (83) ima ni no -mim tavasao
house of class-2 sg. servant
your servant's house
- (84) venatu ni Leban
daughter of L.
Leban's daughter
- (85) venatu i takuni
daughter of man
daughter of a man
- (86) ra'u i 'au
leaf of tree
leaf of a tree

In (83), *tavasao* is certainly not an inalienable noun, as the use of the possessive classifier *no-* indicates. It does however have definite reference.

Ray's account of Kwamera, South Vanuatu, includes the following examples of *i*-phrases:

- (87) numa i nei
leaf of tree
leaf of a tree
- (88) rer i yerama
heart of man
heart of man
- (89) me i fage
handle of shooter
bow

Note that Ray's glosses are not inconsistent with a non-specific interpretation of the genitive noun.

6. RELICS

Many Oceanic languages which have no reflex of *qi in the possessive system, retain as lexical items a number of compound words which contain a linker *i*. Many of these items are idiomatic, metaphorical, or obscure.

In Tongan, in addition to the many compounds with *mata'i-*, we find the following:

- (90) matala 'i akau
flower of tree
flower
- (91) kapakau 'i kava
?wing of ?
variety of yam
- (92) tefito 'i ngāue
base of work
principal work
- (93) tupu'a 'i lao
ancient of law
ancient law
- (94) ko e mata'i hele
foc. art. class. knife
the knife-blade
- cf. ko e mata 'o e hele
of
the blade of the knife

- (95) ko e ngako 'i puaka cf. ko e ngako 'o e puaka
 foc. art. fat pig of
 the lard the fat of the pig

Examples (94) and (95), from Churchward (1953:249), show a contrast between a lexicalised genitive phrase with non-specific possessor introduced by 'i, and the regular possessive structure in which a specific possessor noun is introduced by 'o.

Fijian is also a rich source of these survivals. The productive pattern for genitive phrases in Standard Fijian is NP-ni-NP, and Churchward comments (1941:35) that i before common nouns in the genitive phrase is found only after vula month, as in na vula i katakata the hot months. There are however numerous other forms with i:

- | | |
|--|--|
| (96) na mua i liu
art. front of forward
the bows of a ship | (97) vua i cake
wind of upwards
East wind |
| (98) nātu i cake
line of upwards
the East, windward islands | (99) gauna i liliwa
time of cold
cold season |
| (100) vasu i taukei
moBroSo of indigenous
mother's brother's son, born in mother's village | |
| (101) vatavata i ra lako
? of pers.art. fly
riddle | |

Standard Fijian also has a number of compounds with mata, most having ni but some i as the linking particle. There is no suggestion of the classifier use of mata, except perhaps in the phrases liu mata i rua, mata i tolu go ahead in two ranks, or in three ranks. However the semantic connection with eye or point is usually apparent:

- | | |
|---|--|
| (102) mata i lalai
mata i lelevu
mata i taliga
mataitūtū
matādravu, mata ni dravu | fine-textured
coarse-textured
large axe; hammerhead shark
people of the same yavusa (clan)
fireplace |
|---|--|

The relic which has caused most comment is the so-called independent noun suffix of North Vanuatu languages. This suffix is attached to inalienable nouns when they occur as free forms rather than bound to a suffixed pronoun:

- (103) MTS qatu-i head na qatu-na his head
 sasa-i name na sasa-na his name

It seems likely, on the basis of evidence from the other languages we have examined that this suffix formerly linked two nouns in a 'genitive relation'.

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). There is less evidence for the suggestion made in the same paragraph, that this independent noun suffix originally occurred before personal noun possessors.

While it is the case that in a few North and Central Vanuatu languages, including Aoba (Codrington 1885:422) and Lonwolwol (Paton 1971:34), the head noun does not carry the possessive suffix if the possessor is a personal name, nevertheless the weight of evidence indicates that the usual possessive structure with personal name possessors was the noun stem plus possessive pronominal suffix, with or without a personal article before the proper noun. The following examples from diverse languages justify a POC reconstruction of this possessive structure:

(104)	Tigak:	tang lingi-na i Gamsa	<i>Gamsa's voice</i>
	Roviana:	tasi-na e lone	<i>John's brother</i>
	Bugotu:	toga-ña Mere	<i>Mary's sister</i>
	Kwara'ae:	'ae-na sa Lysias	<i>Lysias' foot</i>
	Merlav:	na lima-n Wenag	<i>Wenag's hand</i>
	Paama:	hei-n Meri	<i>Mary's head</i>
	Eromanga:	iteme-n Simon	<i>Simon's father</i>

The independent suffix derives from a very different kind of phrase. Codrington's description is significant:

The Independent Substantives are so called because these are names of parts, members, things in relation to something which possesses or includes them, but are by this special termination shown to be in thought and in grammar free for the time from this dependence. Thus an eye is in the true form *mata* ... and when any one's eye is spoken of ... this is the stem to which the suffix is applied; *namatana* his eye. But if an eye is spoken of independently of any person, or any organization, the word assumes the termination *i* and becomes *matai* not *mata*.

(1885:260)

We have here a rather well-concealed case of the specific - non-specific contrast. Deck, writing of Kwara'ae, uses wording that is strikingly similar:

When the part of the body is used in the abstract, without reference to any particular person, the possessive suffix is dropped and the genitive preposition 'i is used: 'aena wae the leg of a particular man in view, but 'ae 'i wae, the leg (of man).

(1933:11)

In Mota the expression of non-specific possession has gone a step further, with the dropping of the second noun phrase.²

7. CONCLUSIONS

7.1 The function of POC *qi

I now look at genitive phrases in Arosi, where the two structures which have concerned us exist conveniently side by side. Phrases with specific NP possessors follow the common pattern for South-East Solomonic languages: an inalienable head noun carries the possessive suffix, as in (105) and (106); an alienable head noun is followed by the possessive classifier *a-* plus the appropriate possessive suffix, as in (107) and (108). In both cases the noun-marker or article *i*, termed 'non focus' by Capell (1971:61), precedes the genitive noun:

- (105) ia gare-na i noni
 art. child-suff.art. man
the man's child
- (106) na uwa -na i noni
 art. foot-suff. art. man
the man's foot
- (107) na ruma ana i noni
 art. house class-suff. art. man
the man's house
- (108) mwaeraha adaau i Wango
 chief class-suff.3pl. art W.
the chief of the Wango people

The genitive prepositions *i* and *ni* also occur in Arosi, with the same connotations noted above for other languages:

- (109) ruma ni maho i ngau
 house of thing of eat
storehouse
- (110) nunu i 'abe
 shape of body
carved image
- (111) gare i noni
 child of man
human child
- (112) gare i maa
 child of eye
pupil
- (113) na hereho i ngau 'ana i gare
 art. thing of eat food-class.-suff. art. child
the child's food

Although in Arosi the reanalysis has not taken place, these examples, in particular (113) and the contrasting pair (105) and (111), suggest conditions under which a noun marker or article *i* could be reinterpreted as a possessive marker. This is what I suggest has happened in the four languages considered in section 3, in which *i* occurs as a genitive preposition in phrases with personal noun possessors, which are, of course, highly *specific*.

In spite of the tantalising agreement in the use of *i* meaning *of* before PNPs in Lifu, Standard Fijian, Tigak and Tolai, and the fact that they are not geographically adjacent, it seems wisest to attribute the similarity to parallel development. In each case the personal article *i* has been reinterpreted as a genitive *i*. The Arosi examples given above show that in terms of surface ambiguity this is a well-motivated change. Significantly, none of these languages appears to make use of personal articles. Secondly, it appears that the older *i*-phrases ceased to be a productive part of the grammar, although in Fijian at least numerous examples of these remain, in phrases of undoubted antiquity. Some curious features of the Tolai genitive system can perhaps be attributed to traces of the formerly productive *i* genitives, for example the including of 'canoe' in the class of nouns which select *i*:

- (114) TOL a lua i ra oaga
 art. front of art. canoe
 the bow of the canoe

The evidence indicates that an early stage of Oceanic had a genitive particle *qi, which indicated non-specific or generic possession of inalienable nouns: entities which are a part of, or composed of, or the product of, or a particular kind of some other thing. We can plausibly add, "unit of a series or class". This type of phrase is a frequent vehicle for metaphor or ritual terms:

- (115) ARS gare i maa pupil
 ruma i bao house in which dead body of child is
 house child preserved
 PAA vola i tas skull
 container sea
 FIJ ulu i meri condolence gift
 head ?

Nominal compounds existed alongside i-phrases in most languages, and the availability of this alternative form with the same semantic structure and order of elements probably explains why i has disappeared from so many languages.

7.2 The *qi - *ni contrast

We can now comment on the contrast between *qi and *ni. This is well documented in Kwara'ae and Arosi. In Lonwolwol, it will be recalled, *ni is reflected as ne, and traces of *qi survive only in certain morpho-phonemic patterns, so that in this language the contrast is between ne and Ø. These witnesses agree in that reflexes of *ni occur in genitive phrases of which the head noun is not 'suffix-taking', i.e. is alienable, and reflexes of *qi in phrases of which the head noun is inalienable. Current subgrouping theory considers South-East Solomonic and North and Central Vanuatu to be different primary sub-groups of Oceanic, so on this basis we may reconstruct the *qi - *ni contrast for POC.

It is obvious that the kinds of relationship expressed by ni-phrases (e.g. purpose, location, occupation, attribute), are sufficiently like those expressed by i-phrases to make coalescence of the two structures a likely development. As we saw earlier, ni is the productive genitive particle of Standard Fijian, i surviving only in some archaic forms. Evidence that ni has generalised the functions of *ni and *qi is provided by the non-contrasting variation in the following pairs:

- (116) gauna i liliwa cf. gauna ni leqa
 time of cold time of distress
 cold season time of shortage
 (117) ulu i meri cf. ulu ni vanua
 head of ? head of land
 condolence gift headland
 (118) matādravu, mata ni dravu
 eye ash
 hearth

Interestingly, both non-specific and specific possession may be expressed by the structure NP-ni-NP, suggesting a tendency for ni to invade the full range of possessive expressions:

- (119) na ulu ni tamata
 art. head of person
 human heads
- na ulu -na na tamata oqō / na ulu ni tamata oqō
 art. head-suff. art. person this
 this person's head

7.3 PAN *qi?

At present, we can only speculate as to whether the contrast between *qi and *ni existed at an earlier stage of Austronesian, or was an innovation of POC. (See Reid 1981 for possible reflexes in Philippine languages of a PAN genitive determiner *i.) It has been shown by Blust (1974) that idiomatic genitive phrases existed in PAN and have survived in daughter languages with the same meanings. To the Toba Batak and Fijian comparisons containing ni presented by him we can now add some Oceanic reflexes with *qi as the genitive particle:

- | | | | | |
|-------|-----|---------------|--------------|--------------------------|
| (120) | TOB | mata ni susu | mata ni ihan | mata ni anin |
| | FIJ | mata ni sucu | mata ni ika | mata ni cagi |
| | TON | mata'i huhu | mata'i ika | matā matangi |
| | LON | mete susu | | SAA maa i deni |
| | | <i>nipple</i> | <i>sore</i> | <i>direction of wind</i> |

Note also:

- | | | | |
|-------|-----|--------------|-------------------------------|
| (121) | TOB | mata ni uvi | <i>yam tops</i> |
| | FIJ | mata ni dalo | <i>taro tops</i> |
| | TON | mata'i talo | <i>taro tops</i> |
| | PAA | mete i tan | <i>mound over planted yam</i> |
| | SAA | maa i aro | <i>taro tops</i> |

Blust assumes on the basis of the Fijian and Toba Batak forms that NP-ni-NP is reconstructable as the PAN form. This is not a logical necessity, as the other reflexes show. NP-i-NP forms in Fijian are quite evidently archaic. As Andrew Pawley has pointed out (personal communication), NP-ni-NP is a productive pattern in Fijian and speakers would be fully capable of using ni where once there had been an i or qi. Tongan, on the other hand, not only has no genitive ni, but also lacks a fully productive pattern with 'i. Hence the Tongan forms offer reliable evidence of archaism.

If *qi were an innovation of Oceanic, we would have to assume that the generalised NP-ni-NP pattern of Fijian was the result of two successive changes. Firstly, at an early stage of Oceanic, an innovation *qi replaces PAN *ni in certain environments. Secondly, in Fijian (and a few other languages such as Nggela) *ni supplants *qi in all but a few phrases. On the other hand, if both *ni and *qi were used in PAN, then one and the same change has taken place in Fijian and Toba Batak: *ni has generalised the functions of *qi and *ni.

NOTES

1. I wish to thank Andrew Pawley for innumerable helpful comments and criticisms, for bringing much relevant data to my attention, and for helping to collect the data for Bugotu and Standard Fijian. I also wish to thank Ross Clark and Frank Lichtenberk for making most helpful comments on the final draft. The following informants supplied data from their languages: Alan Marat (Tolai), Trixie Legua (Bugotu), and Timoci Sayaba (Standard Fijian).
2. In Motlav, Volow and Lakon, the independent noun suffix has a variant -n. The examples given by Codrington suggest no consistent semantic or phonological motivation for this variation. For example, in Motlav *ban hand* takes the -ge variant of the suffix, whereas in Lakon *pane hand* takes -n. The genitive particle *ni* or the third person singular possessive suffix, which has the shape -n in all three languages, are possible sources of this form.

APPENDIX

This section lists a selection of genitive phrases incorporating classifiers from a number of Oceanic languages. No attempt is made to separate literal from metaphorical or idiomatic expressions.

Arosi

rawa i niu	<i>100,000 coconuts</i>
e siha abe i manu	<i>how many (bodies of) birds?</i>
kora i daro	<i>ankle bone</i>
kora i hau	<i>skull (lit. round thing of stone)</i>
kora i uhi	<i>Haliotis shell</i>
kora i tete	<i>round end of a club</i>
hua i i'a	<i>a fish</i>
hua i hau	<i>a stone</i>

Kwaio (see also section 4.2)

fe'e ba'u	<i>a banana</i>
fe'e golu	<i>heart</i>
olu fe'e wiki	<i>three weeks</i>
k ^w ali'i niu	<i>a single coconut</i>
fu'u'i ba'u	<i>a bunch of bananas</i>
fuu'i wane	<i>a group of men</i>
tofu'i boo	<i>a piece of pork</i>
g ^w e'e i'a	<i>a whole fish</i>
g ^w e'e alo	<i>taro corm</i>

Lau

maae fera	<i>country, village</i>
maae dangi	<i>a day</i>
maae rodo	<i>night, darkness</i>
fai maae ooru	<i>the four winds</i>
maae fote	<i>a paddle</i>
maae furai	<i>a net</i>

Fijian (see also section 6)

mata i tolu	<i>in three ranks</i>
mata i rua	<i>in two ranks</i>
mata i lalai	<i>fine-textured</i>
mata i lelevu	<i>coarse-textured</i>
mata i taliga	<i>large axe; hammerhead shark</i>

Tongan

matāfanga	}	<i>beach</i>
matātahi		
matā'one'one		
matāfefine		<i>fine-looking woman</i>
matātangata		<i>fine-looking man</i>
matā'italo		<i>taro tops for planting</i>
matā'ifika		<i>numeral</i>
mata'i tohi		<i>letter of alphabet</i>
mata'ihuhu		<i>nipple</i>
mata'imoho		<i>seeds of moho vine</i>
mata'ipeni		<i>pen nib</i>
mata'itafa		<i>incision, wound left by incision</i>
mata'iuha		<i>raindrop</i>
mata'ipoto		<i>clever-looking</i>
mata'itofe		<i>pearl of oyster</i>
mata'itevolo		<i>devil-faced</i>
fo'ia'akau		<i>fruit; pill</i>
fo'i 'ao		<i>genitals</i>
lau'i tohi		<i>a page of a book</i>
lau'i pate		<i>the blade of a bat</i>
lau'i matala		<i>a petal</i>
mata'ikoloa		<i>foremost in one's estimation</i>

Samoan

mata'ioloa	<i>the best item of goods</i>
mata'ifale	<i>incest</i>
mata'inema	<i>numeral</i>
mata'itusi	<i>letter of alphabet</i>
mata'isiva	<i>the best dancer</i>
mataitagata	<i>fine-looking man</i>
mataitōga	<i>the best tōga (fine mat) of the lot</i>
matāutu	<i>point of land</i>
matāfaga	<i>beach</i>
matāmatagi	<i>direction of wind</i>
matāmeli	<i>drop of honey</i>
matāfale	<i>gable of house</i>
matāgāluēga	<i>section, department (as in Treasury Department)</i>

Samoan

mata'upu	<i>subject, theme</i>
fua'i'upu	<i>sentence, remark</i>
fua'ifa'i	<i>a single banana</i>
fua'itau	<i>a line of a song</i>
fuaivai	<i>a single water bottle</i>

Nukuoro

madaaduge	<i>top of a duge</i>
madaa gai	<i>first fruits</i>
madaa hadu	<i>metal grater</i>
madaa moni	<i>proW of canoe</i>
madaa sele	<i>loop of noose</i>
madaa ua	<i>drop of rain</i>
hua	<i>classifier (by tens) for fruit</i>
hua i gadea	<i>side of canoe away from outrigger</i>

Kapingamarangi

madaa dogi	<i>plane blade</i>
madaa fale	<i>end of house</i>
madaa li	<i>nipple</i>
madaa hua	<i>narrowing of hull at ends of canoe</i>

Luangiua

makaa ko'i	<i>raindrop</i>
makaa lime	<i>fingers</i>
makaa vae	<i>toes</i>
makaa va'a	<i>fore or aft of canoe</i>
makaa 'uu	<i>nipple</i>

Rennell and Bellona

mataa kape	<i>top of taro</i>
mataa baka	<i>canoe prow</i>
mataa uu	<i>nipple</i>
mataa 'ua	<i>raindrop</i>
mataa	<i>ten small fish (classifier)</i>
huaa biti	<i>beads</i>
huaa 'umanga	<i>planted root crops</i>
huaa ga'akau	<i>fruits</i>

Futuna-Aniwa

matā tui	<i>100 or more fish</i>
matā puke	<i>top of yam mound</i>
foivai	<i>calabash</i>
foimata	<i>eye, face</i>
foifata	<i>crop of fruit</i>

Some semantic sets

Canoe parts

BUG	ulu i hore	}	<i>bow</i>
MTA	qat i aka		
TOL	lua i ra oaga		
FIJ	mua i liu		
	ucu i mua		<i>stays from the masthead</i>
NUK	madaa moni		<i>bow</i>
	hua i gadea		<i>side of canoe away from outrigger</i>
KAP	madaa hua		<i>narrowing of hull at ends of canoe</i>
LUA	makaa va'a	}	<i>bow</i>
REN	mataa baka		
HAW	kama i kahuli wa'a		

Eye

ARS	gare i maa	}	<i>pupil of the eye</i>
TON	tama'i mata		
SAM	tama'i mata		
TOK	io i mata		

Wind

LAU	fai maae oru		<i>the four winds</i>
SAA	maa i deni		<i>a wind</i>
FIJ	mata ni cagi		<i>direction of wind</i>
	vua i rā		<i>west-north-west wind</i>
	vua i cake		<i>east wind</i>
TON	matāmatangi	}	<i>direction of wind</i>
SAM	matāmatagi		

Yam, taro

SAA	maa i aro		<i>taro top for planting</i>
PAA	mate i tan		<i>mound over yam plant</i>
FIJ	mata ni dalo	}	<i>taro top for planting</i>
TON	mata i talo		
SAM	matā talo		
REN	mataa kape		
FUT	matā puke		

BIBLIOGRAPHY

BEAUMONT, Clive

- 1979 *The Tigak language of New Ireland*. PL, B-58.

BLUST, Robert

- 1974 Proto-Austronesian syntax: the first step. *Oceanic Linguistics* 13:1-15.
- 1977 The Proto-Austronesian pronouns and Proto-Austronesian: a preliminary report. *Working Papers in Linguistics, University of Hawaii* 9/2:1-15.

CAPELL, A.

- 1930 The language of Inakona, Guadalcanal, Solomon Islands. *Journal of the Polynesian Society* 39:113-136.
- 1941 *A new Fijian dictionary*. Sydney: Australasian Medical Publishing Company.
- 1971 *Arosi grammar*. PL, B-20.

CARROLL, Vern and Tobias SOULIK

- 1973 *Nukuoro lexicon*. Honolulu: University of Hawaii Press.

CHURCHWARD, C.M.

- 1941 *A new Fijian grammar*. Sydney: Australasian Medical Publishing Company.
- 1953 *Tongan grammar*. London: Oxford University Press.
- 1959 *Tongan dictionary*. London: Oxford University Press.

CODRINGTON, R.H.

- 1885 *The Melanesian languages*. 2nd edn, 1974. Amsterdam: Philo Press.

CODRINGTON, R.H. and J. PALMER

- 1896 *A dictionary of the language of Mota*. London: Society for Promoting Christian Knowledge.

CROWLEY, Terry

- 1982 *The Paamese language of Vanuatu*. PL, B-87.

DECK, N.C.

- 1933 *A grammar of the language spoken by the Kwara'ae people of Mala, British Solomon Islands*. Polynesian Reprint no.5. Wellington.

ELBERT, Samuel H.

- 1975 *Dictionary of the language of Rennell and Bellona, Part 1: Rennellese and Bellonese to English*. Copenhagen: National Museum of Denmark.

FOLEY, W.A.

- 1976 Comparative syntax in Austronesian. Ph.D. thesis, University of California at Berkeley.

FOX, C.E.

1970 *Arosi-English dictionary*. PL, C-11.

FRANKLIN, Karl J. and Harland B. KERR

1962 *Tolai language course*. 2nd edn 1968. Port Moresby: DIES.

IVENS, Walter G.

1921 *Grammar and vocabulary of the Lau language, Solomon Islands*.
Washington: Carnegie Institute, Publication 300.

1927 A study of the Oroha language, Mala, Solomon Islands. *Bulletin of the School of Oriental and African Studies*, 4:587-610.

1929a *A dictionary of the language of Sa'a (Mala) and Ulawa, South-East Solomon Islands*. Oxford University Press.

1929b A study of the language of Marau Sound, Guadalcanal, Solomon Islands. *BSOAS* 5:345-358.

1931 A grammar of the language of Kwara'ae, North Mala, Solomon Islands. *BSOAS* 6:679-700.

1933 A grammar of the language of Bugotu, Ysabel Island, Solomon Islands. *BSOAS* 7:141-177.

1937 A grammar of the language of Florida, British Solomon Islands. *BSOAS* 8:1075-1110.

1940 A grammar of the language of Lobaha, Lepers' Island, New Hebrides. *BSOAS* 10:345-364.

1940 A grammar of the language of Lotoro, Maewo, New Hebrides. *BSOAS* 10:679-698.

KEESING, Roger

1975 *Kwaio dictionary*. PL, C-35.

LANYON-ORGILL, Peter A.

1962 *A dictionary of the Raluana language*. Victoria, British Columbia: The author.

LICHTENBERK, Frantisek

1985 Possessive constructions in Oceanic languages and in Proto-Oceanic. In this volume, 93-140.

MILNER, G.B.

1966 *Samoan dictionary*. Oxford University Press.

MOSEL, Ulrike

1977 *Tolai texts*. *Kivung* 10:1-2.

1980 *Tolai and Tok Pisin: the influence of the substratum on the development of New Guinea Pidgin*. PL, B-73.

PATON, W.F.

1971 *Ambrym (Lonwolwol) grammar*. PL, B-19.

1973 *Ambrym (Lonwolwol) dictionary*. PL, C-21.

PAWLEY, Andrew

- 1972 On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history*, vol.3, 1-142. Honolulu: Bernice P. Bishop Museum.
- 1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103-188.

PRATT, G.

- 1911 *Grammar and dictionary of the Samoan language*. Samoa: Malua Printing Press.

RAY, Sydney H.

- 1926 *A comparative study of the Melanesian island languages*. Cambridge University Press.

REID, Lawrence A.

- 1981 Proto-Austronesian genitive determiners. In Andrew Gonzalez FSC and David Thomas, eds *Linguistics across continents: Studies in honor of Richard S. Pittman*. Manila: Summer Institute of Linguistics and Linguistic Society of the Philippines.

SCHÜTZ, Albert J.

- 1969 *Nguna grammar*. *Oceanic Linguistics* Special Publication No.5. Honolulu: University of Hawaii Press.

RAPID LEXICAL CHANGE AND ABERRANT MELANESIAN LANGUAGES: SENSENG AND ITS NEIGHBOURS

Ann Chowning

1. INTRODUCTION

Almost twenty years ago, arguing against blind reliance on lexicostatistics, Grace suggested that some of those Melanesian languages which "show very few identifiable cognates with other AN languages ... are characterized by very low rates of retention (in the lexicostatistical sense), by complicated sound changes, or by both" (1964:366). More recently, he has pointed out that the situation is probably much more complex, with "many different kinds of aberrancy" (1981: 256). I agree with that suggestion, but in this paper wish to deal primarily with one of the possibilities he raised in his earlier paper. We now have examples of languages in which the sound changes do indeed make it difficult for a casual observer to recognise AN forms; see, for example, Lynch's work (1975) on Lenakel of Vanuatu. An example of rapid lexical change within this century has been documented by Lithgow (1973), who touches very briefly on some of the causes, including the modern phenomenon of influence from a mission language, but does not discuss any of them in detail. Meanwhile, however, several writers have discussed one particular cause that is known (or sometimes assumed) to have affected retention rates in some Oceanic languages and that may have done so in many more. This results from temporary or permanent tabus on the use of certain personal names, which Clark called a "widespread Polynesian linguistic practice", and to which he attributed depression of cognate percentages in several Polynesian languages, particularly Tuamotuan and Tahitian (1979:265-266). In an article published earlier, describing how name avoidance operates in Kwaio of Malaita, Keesing and Fifi'i (1969:155) suggested that perhaps it "was characteristic of some or all early Austronesian speakers in the Pacific", and went on to mention "the possibility that the process has significantly accelerated vocabulary differentiation between genetically related languages", causing various problems for those attempting to classify them. For example, "such tabooing could create a spurious impression of long divergence or skew datings, or in some cases even hide genetic connections." More recently, Simons has discussed the effects of name tabu in two regions, Santa Cruz as well as Malaita. He mentions a problem not discussed in detail by Keesing and Fifi'i, the likelihood that languages so affected will develop many true synonyms, influencing not only cognacy rates (if all synonyms are not recorded by an investigator) but the establishment of dialect-language boundaries (since substitute forms are so often borrowed) (Simons 1982:162-167).

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 169-198. *Pacific Linguistics*, C-88, 1985.

© Ann Chowning

169

In the same paper, noting that in the discussion of name tabus in *The Golden Bough* most of the examples (so it seems) were from AN languages, and surveying the reported incidence of different types of word tabus reported for AN and NAN-speakers over a wide area, Simons concludes that: "For the Proto-Austronesian speech community, there was ... a name taboo between parents-in-law and children-in-law"; that five other kinds of 'name taboos' can be reconstructed for "Proto-Oceanic or at least major subgroups of it"; and that "All of these name taboos extended to a taboo on the common use of words occurring in the name" (1982:187). My reading of Frazer does not support Simons' suggestion that most of Frazer's examples came from AN languages, and Simons' own survey makes it clear that most types of tabus on personal names and other words are well recorded for NAN languages and languages outside Oceania. In particular, as anthropologists have noted since the 19th century (Tylor 1870:146-150), name tabus are found throughout the world, and have been shown to affect everyday vocabulary in societies as far from Oceania as the Caribbean and Imperial China (Jameson 1956:782; Metraux 1956:783). If name tabus are typically Oceanic, they are hardly peculiar to that region. Nevertheless, it may be that in many parts of the Pacific local conditions – geographic, demographic, and cultural – made it particularly likely that they could so affect the general lexicon (rather than individual usage) as to make the language as a whole seem to have split off early from other related languages; seem possibly to have been affected by contact with speakers of very different languages (see Clark 1979:265; Grace 1981:256); or seem possibly not to be AN at all. However much their vocabulary may have been affected by word tabus, no one appears to have denied that Tuamotuan, Tahitian, and Kwaio are AN languages, but a number of linguists (Loukotka 1957, Capell 1971, and most recently Blust 1981) have been reluctant to grant that status to the languages that I am about to discuss.¹ I am arguing first that they are AN but have undergone an exceptional amount of lexical replacement; second that name tabus may have been the reason for rapid lexical change in the past, as they are today; and third, along with Simons, that other aberrant Melanesian, as well as Polynesian, languages may well have lost AN content for the same reason.

2. PASISMANUA LANGUAGES

The languages concerned are those in what I have labelled the Pasismanua division of the Whiteman group, located around but mostly south of the Whiteman Range in south-west New Britain. From west of the Alimbit to east of the entire Whiteman range, they consist of Miu, Kaulong, Sengseng (Asengseng)², Karore³, and what Johnston calls Psohoh, a dialect chain extending from Getmatta in the south to Bao in the north (see maps in Chowning 1976 and Johnston 1980b). Johnston has recently (1983) added Uvol to this subgroup, which corresponds to his Western Whiteman, but I find the evidence for its inclusion unacceptable. I shall therefore confine myself to the languages just named, but concentrate on those for which I have the most information, Sengseng (studied by myself) and Kaulong (studied by Goodale and more recently by C. and L. Throop).⁴ Most of my data are taken from three Sengseng villages in which I lived, and from the two Kaulong villages in which Goodale lived, supplemented by information from other parts of the area.

Sengseng, distributed along either side of the Andru River east of Kandrian, is flanked by Kaulong to the east and Karore (another Pasismanua language) to the west, with uninhabited bush behind it, but Arawe languages are spoken on

small islands located just off the coast at the eastern end of the Sengseng area. By contrast, Kaulong is flanked on the south and west by Arawe languages which are also spoken on the mainland of New Britain (see maps in Chowning 1969 and 1976). Arawe languages were, then, the only other ones in direct contact with Pasismanua languages other than Miu, the westernmost language, in the period immediately preceding European contact, although it is probably safe to say that trade brought many of the western Pasismanua-speakers into contact with speakers of Lamogai languages (see Chowning 1978a).

Throughout south-west New Britain, groups in the interior traditionally lived in very small settlements scattered widely over the countryside, a fact that explains their comparatively late contact with the outside world. An effect of government patrols and pressures from missionaries and others such as cult leaders has been to encourage (or force) people to build larger consolidated villages and also to move nearer the coast. These shifts have also made it difficult to ascertain the numbers of speakers of each language because so many villages now have mixed populations. Nevertheless, it can be said that numbers ranged from about 3,000 Kaulong speakers down to fewer than 400 for Miu (and undoubtedly still fewer for Karore), in 1980. At the time of this census predominantly Sengseng villages, some of which contained a considerable number of Karore-speakers, had a total population of 865.

Pacification came late to the interior of the Pasismanua. When Goodale and I began fieldwork there in 1962, the interior Kaulong and Sengseng villages had only been brought under government control less than a decade previously, at various times (for different villages) during the 1950s, and Miu was still uncontrolled. With pacification, unmarried men began to go out to work, most commonly only to coastal plantations in neighbouring, usually Arawe-speaking regions, in order to earn enough money to buy foreign goods, particularly steel tools. A few, however, went as far as Rabaul or Manus, lured by access to cheap supplies of goldlip pearl shells, the major form of wealth in the Pasismanua. Only these latter, amounting to just one or two men in each tiny village, learned fluent Pidgin. The other workers picked up a little Pidgin and other words which they thought were Pidgin but which are unknown outside that region, and which presumably come from other New Britain languages (cf. Chowning 1983).

Nearer the coast, pacification and wage labour had begun much earlier, and villages very near the coast and the government station of Kandrian (Moewehafen) also had access to village schools (which as of 1981 had still not been established deep in the interior). In the 1960s, many Pidgin and pseudo-Pidgin words and phrases had entered everyday vocabulary, though some people were unaware of their source,⁵ but in the interior only a few young men claimed to speak the language. Children of both sexes were gradually picking up a smattering of Pidgin from the young men, but adult women and men who had not been out to work could not engage in or understand extended discourse in Pidgin. The few who were not monoglot understood or, more rarely, spoke one of the neighbouring south-west New Britain languages, either because of living in a border region or as a consequence of a marriage between speakers of different languages. Descendants of such marriages often maintained connections with foreign kin through extended visits, in the course of which they learned the other languages (whereas the inmarrying spouses I knew did not consciously maintain their own languages, however much they may have served as a source of the innovations to be described below).

Sengseng, Kaulong, and the other Pasismanua languages are very closely related,⁶ to such an extent that it can be argued that they are only dialects of the same language, and indeed the Throops have suggested this (1980:228), but I have used the criterion of mutual intelligibility (see Pawley 1981:271), as based on my own experience and statements from local people, and on these grounds have called them different languages. People who live in border areas usually learn to understand the neighbouring language, and some border villages contain many speakers of one of the other languages, either because of inter-marriage or recent migration. (In particular, the Karore-speaking region to the east has been heavily infiltrated by migrants from interior Sengseng villages.) The local people nevertheless firmly identify themselves and each other as speakers of a particular language and quite often of a dialect within it. Many Sengseng and Kaulong specify the differences between their languages in terms of a few common words and phrases, notably K. e mo, S. a moi *you come*; K. maŋ, S. masaŋ *men's house*; K. e-gin, S. e-ki *bird*; K. yok, S. tuwo *father* (address term).⁷ That the people who mention these differences have only a superficial knowledge of the other language is shown by their usual failure to note or be aware of semantic differences such as the fact that tuwo is only used for the true *father*, with a special term for the *father's brother*, whereas yok encompasses both, or of other peculiarities such as the fact that the S. word for *bird* has fallen together with the word for *water* while the K. one has not.

Because Kaulong is spoken by a much larger number of people than Sengseng, and the government station, trade stores, and mission station can only be reached through Kaulong territory, it is typically the Sengseng who have a working knowledge of the other language. In border areas and mixed villages, it seems that most people have at least a passive understanding of the other language, as well as some confusion about which words or other forms such as prepositional endings are properly assignable to which language. The same confusion exists regarding dialectical differences. The Sengseng recognise two dialects, those of the 'beach' and the 'bush'. Informants from the village near the coast in which I lived in 1980-1981 described the difference wholly in terms of vocabulary (while quite often telling me that words I had learned in the bush were really Kaulong forms), and never mentioned minor differences in pronunciation, such as a tendency near the beach to voice or trill word-final and preconsonantal /t/ and to produce a bilabial fricative before /u/ rather than the semi-vowel /w/ that appears in the interior. Goodale's Kaulong-speakers from Angelek mentioned four or possibly five dialects of Kaulong, characterised by vocabulary and in at least one case pronunciation. Both Sengseng and Kaulong distinguish this kind of dialectical variation from the differences between their two languages. (On the other hand, although they consider it a separate language, Sengseng-speakers typically characterise Karore only as substituting /r/ for Sengseng /h/.)

3. INITIAL IMPRESSIONS OF THE LANGUAGES

In order to justify the inclusion of this paper in this symposium, it is first necessary to indicate why I think that these languages are AN. I first entered the area in 1962 in search of two adjacent groups of AN-speakers who had not yet been converted to Christianity, since the plan was that Goodale and I would carry out a comparative study. I already spoke two AN languages, Lakalai (West Nakanai) of West New Britain and Molima (Morima) of Ferguson

Island, and apart from my general interest in related languages, thought that another AN one would be easier to learn than most of the reputedly difficult NAN languages of the New Guinea area. In addition, of course, I hoped that my knowledge of Lakalai and Molima would help me with a third language. (I should add that between my third and fourth trips to Sengseng I undertook study of another AN language, Kove of West New Britain, and some of my later comparisons were drawn with it.)

I travelled through Kaulong-speaking territory, where Goodale remained, and on to Sengseng, spending eight weeks in the area and concentrating on collecting linguistic material to be studied before undertaking a second longer trip. Despite the small amount of overt AN forms in the basic vocabulary, I decided almost at once that these were AN languages, but since I have obviously failed to convince a number of other linguists, I need to set out my reasoning in some detail. I was not particularly concerned by the low number of cognates; although Lakalai (like Kove) is an 'exemplary' language (Grace 1971:345; see Chowning 1973), Molima is not. Sengseng, like Kaulong, did contain a number of obvious AN forms in basic vocabulary, and various others looked possible. The obvious ones included some body parts (*mata-eye*;⁸ *mamai-tongue*); verbs (*num drink*; *sus suck milk*; *kel dig*); pronouns (*ita we inc.*; *i mata-n its eye*); the principal connective (*ma and*) and the productive causative prefix *pa-*. Of the possibilities, some had an unexpected vowel or consonant — e.g. *moi, me come, hither*; *kliŋa-ear*;⁹ *sinaŋ sun*; *sihit sew*; *pima we exc.* — but still looked very likely, while others were more uncertain either because so much of the proto-form was missing, as with *lit skin*, or because the sound shifts seemed particularly unlikely, as with *umat stone* and *e-mut louse* (where *e-* is an article.) (Many of these will be discussed below.) I was particularly struck by the fact that although many of the forms looked AN, they often did not closely resemble Lakalai forms, even though I had some reason to suspect borrowing between the subgroups to which Lakalai and Sengseng belonged, nor did they closely resemble the forms in other nearby languages, with one or two exceptions to be mentioned below. For example, the Lakalai word for *drink is liu*, and although it is derived, like the Sengseng one, from PAN **inum*, obviously a different history is involved. (Here Sengseng resembles Molima, which has *numa*; in both languages, but not in Lakalai, the third person singular nominative pronoun is *i*, and presumably the initial vowel of the verb was assimilated to it.)¹⁰

The question of whether Sengseng could have acquired its AN component purely by borrowing from other New Britain AN languages will be dealt with in much more detail below, but my impression then as now was that it could not. Furthermore, Sengseng grammar also struck me as fundamentally AN. First, there was no sign of the elaborate system of noun classes reported for some NAN languages of East New Britain, or for Mungen, which is often stated to have been influenced by NAN (see Chowning 1978b:1136). The whole pronominal system, and in the case of the singular and plural, many of the actual forms, fitted well with what I knew of other OC languages, with one major exception, the presence of sex differentiation in third person singular pronouns referring to human beings. Otherwise, familiar features including an inclusive-exclusive distinction; division into singular, dual, paucal, and plural forms; the verb phrase with a subject marker preceding and a direct object pronoun following; and the use of suffixed possessives with body parts and kinship terms along with a set of separable possessives in other cases (see Table 1). In addition to the S-V-O structure of the verb phrase, I found not only the above-mentioned causative *pa-* but the use of reduplication to indicate ongoing action, futurity/intention marked by a particle *ka* (for first person subjects) or *ko* (for the rest)

preceding the verb¹¹; completion of action indicated by *kut* (from POC *qoti?) following the verb; and reciprocity indicated with a suffixed *-val* (see POC *paRi). The fact that this last is not a prefix was the only unexpected feature; in function, as in shape, it seems like reciprocals in other OC languages (see Chowning 1978b:1174). Verbs are formed in to nouns by suffixing *-ŋin*. I was slow to realise that personal names have prefixes indicating the sex of the person, though the same feature is well known for Tolai. Relativisation is accomplished by devices found in other West New Britain OC languages (Chowning 1978b:1142), but Sengseng lacks some of their other grammatical features, such as infixes, special plural forms, and a suffix indicating transitivity. Nor are there any of the postposed locatives of the kind I was already familiar with from Molima, and was to find in Kove. The only grammatical feature which could be considered NAN is sex gender in pronouns, which is also found in Baining of East New Britain (Wurm 1975:790). Of course, Sengseng, presumably like any other language, has distinctive grammatical features as well, most of which are mentioned in my 1978b paper, but they neither seemed NAN nor made it a difficult language to learn. My problems with the language, apart from those caused by the dearth of people who could translate into Pidgin, derived to a slight degree from phonology and to a greater one from the proliferation of synonyms which will be examined and explained below.

As regards phonology, little difficulty was caused by the consonants apart from occasional failure to detect a final stop.¹² The initial consonant clusters which obviously strike many as NAN, at least superficially, are not difficult to pronounce, in my experience. I did, however, find it hard to differentiate vowels, and to decide how to record them, both in the *e-i* range and in the *o-u* range. Judging from their varying recordings of the same morpheme, other linguists better at phonetics than I have also encountered difficulties. This difficulty extends to languages outside the Pasismanua subgroup. Referring to all the languages of south-west New Britain, and particularly mentioning Combs' inconsistent orthography for Mangsing, Johnston notes that the available word-lists are "highly suspect" in exactly the same range (1983:24). Combs reconstructed seven vowel phonemes for Mangsing (classified by me as related, though distantly, to the Pasismanua languages), and Johnston tentatively did the same for Bebeli, (which he assigns with Mangsing to Eastern Whiteman) and for Psohoh (Johnston 1980b:124,127). That the vowels differ from other Melanesian languages I know is suggested by Sengseng pronunciation of some Pidgin words, such as *bosi* for *pusi cat*. I mention this because sometimes I may postulate that a Sengseng word which I have recorded with /u/ for example, is derived from a proto-form reconstructed with /o/ (as in the *kut- *qoti* example given above) without trying to account for an actual vowel shift; in some of these cases, I do not trust my own recording. I have never achieved a satisfactory phonemicisation of Sengseng vowels. (Here, for convenience, I have arbitrarily reproduced forms as if there were only five vowels.)

Following my initial visit to Sengseng, I spent a further 18 months there carrying out anthropological fieldwork. The data I collected during that time only strengthened my belief that the language is Austronesian.

4. REFLEXES OF POC AND PAN FORMS

Despite problems, I have taken most of the examples that follow from Sengseng rather than neighbouring languages for three reasons. First, I know where I am likely to make mistakes, and am uncertain about some of the other data.

Second, I have much more material on Sengsen. Third, in general Sengsen manifests medial consonants which are not found in Kaulong, and I assume that the Sengsen form is the more conservative one. Also, Sengsen often has a final /-a/ or /-i/ where Kaulong has /Ø/. Examples include S. -lut-, K. -lusibling *opposite sex*; S. ita, K. it *we inc.* (This general shortening of pronouns is perhaps not found in all Kaulong dialects, though it is also reflected in the list in Throop 1980:230.) On the other hand, it must be mentioned that Sengsen differs from related languages to the east and west in having frequently dropped a final -ŋ or -n that appears in many common words, and so is apparently unique in the area in having the words for 'bird' and 'water' fall together (cf. Kaulong e-gin, Karore and Psohoh e-giŋ *bird*; Sengsen e-ki). Other cases in which Sengsen has a shorter form are a few in which Psohoh shows three consonants together, and Sengsen lacks the voiced stop: Psohoh mbrit, Karore mirit, Sengsen mihit *shoulder, wing*.¹³ Finally, in at least one common word — Kaulong ploŋ-, Sengsen loŋ *fall* — Kaulong has an initial consonant which Sengsen lacks. Although almost none of the words in which Sengsen lacks phonemes found in other languages have seemed to be OC (or AN) in origin, these variations will be mentioned when they seem to shed some light on the possible history of forms.

4.1 Reflexes of POC (and PAN) consonants in Sengsen

As I have already mentioned, Sengsen and its neighbours are characterised not only by a paucity of AN forms but by a plethora of synonyms. I hope to show that these two phenomena are related, but before doing so it is necessary to tackle two other questions. The first is whether Sengsen in fact contains a higher AN content than appears at first glance, and the second — and much more difficult — is how much of such content can be attributed to borrowing rather than direct inheritance. Simply because few forms are obviously AN in origin, it is difficult to collect enough cases to establish regular sound shifts, and a paucity of data apart from basic vocabulary in other south-west New Britain languages also makes borrowings hard to identify. What follows is perforce tentative.

An examination of the total lexicon makes a few tendencies clear. As far as I can ascertain, Sengsen has not undergone drastic sound shifts, though it is of course possible that I have failed to detect many AN forms. In saying this, I disagree with Johnston 1983, in that I think that many of the forms that he derives from a single Proto-West New Britain proto-form are not even cognate with each other, much less derivable from one that can be called OC. For example, S. pe-luwok *night* cannot possibly come from *mpoŋi, since pe- is a prefix (translated as Pidgin ples) found in all terms referring to time and weather. Johnston's hypotheses and postulated proto-forms will be discussed further below.

The most common correspondences with the phonemes reconstructed for POC and PAN are shown below.

4.1.1 Examples

Most POC forms have been taken from Lincoln 1979. Where one of the meanings assigned to the POC form agrees with the Sengseng item, I have not given glosses for both. Uncertain cases are discussed in more detail. If a proto-form has been reconstructed only for PAN, it is listed after the POC forms. A question mark indicates that I am uncertain about cognacy; "etc." that various other examples could be supplied.

POC *k usually remains /k/, except in pronouns (see Table 1).

Examples: *kapi *grasp, etc.*; S. e-kap *tongs*; *kaRati, S. kalat, klat *bite, bite through*; *kasi, S. kas *scratch, itch*; *kati, S. kat *bite*; *katemo, *timo, S. katim *native cucumber*; *keli, S. kel *dig*; *kimo, S. kikim *blink*; and various other examples of non-initial *k; *sakil *stamp, S. sak step on*; *suki, S. suk *thrust into*; e-suk *digging stick*.

I suspect S. nekoit *octopus* (POC *kuRita) along with several other nouns (e.g. nepui *paddle*) of being a borrowing because of the unexpected prefix (fossilised article?); in Sengseng, ne/ni is usually an adjectival prefix or relativiser.

S. e-mut (?POC *kutu) is aberrant. In neighbouring languages the word for *louse* is ɲut, a more plausible derivative, and sometimes, S. /ŋ/ is reflected by Kaulong /m/. At best, this is probably a borrowing.

POC *m remains S. /m/.

Examples: *mata, S. mata- *eye*. *maya, S. mamai- *tongue*; *ma, S. ma *and*; *malaque *open space in village*, *malala *cleared ground, S. mla exposed, in the open (etc.)*; *inum, S. num *drink*; *tama, S. tama- *father, etc.*

POC *n remains /n/.

Examples: *nunu, S. e-nu *shadow, reflection*; *pani, S. pan *paint*; *tunu, S. tun *set fire to*; *tina, S. tina (address) *mother, etc.*

POC *ñ is reflected as /n/.

*-ña, S. -n *third p.s. possessive suffix*; *ñami, S. manman (with metathesis) *taste* - see Blust 1977), but as /h/ in one case: *ñamuk, S. humuk *sandfly, mosquito*.

POC *ŋ usually remains /ŋ/.

Examples: *saŋa, S. saŋa- *groin, saŋan fork*; *laŋo, S. e-laŋ *fly*; *yaŋo, S. yaŋ *yellow*; *poŋ *foolish, S. poŋoŋon ignorant of it?*

In two words, however, it is reflected as /n/: *taŋis, S. tinis *weep*; *ŋas *chew, S. nas chew sugarcane* (Blust 1977). The former is a comparatively rare synonym for S. hau *weep, cry out*, but the latter is the usual form. See note 14.

POC *p usually remains /p/.

Examples: *pa-, S. pa- *causative prefix*; *paRi, S. e-pa *stingray*; *puti, S. put *pluck*; *kapi, S. e-kap *tongs, etc.* Sometimes, however, it is reflected as /w/, either an allophone of S. /v/ (a bilabial spirant) or of /u/: *paRi, S. -val *reciprocal*; *pale *house, S. a-val hut* (sloping rather than gabled roof). S. /v/ and /w/ are not recorded

in word-final position. A single example of *p producing S. Ø seems to be S. kau *fine ashes, earth burnt to lime-like consistency, dust* (POC *kapu). This appears as kau in other SWNB languages (see Johnston 1983). If S. umat *stone*, with metathesis, is from *patu, it is presumably a borrowing from one of the languages to the east. For discussion of the cases in which *p may be reflected as /y/, see 4.3.

POC *s is usually reflected as /s/.

Examples: *susu, S. sus *suck milk*, sus(u)- *breast*; *masin *salt*, S. masɨn *salty, tasty*; *sulu, S. e-sul *torch*; *somod *dirty*, S. sumuh *dirt*, etc. The one exception noted is -hi, plural marker and possessive (*sida *they*). For another irregular derivative of this POC proto-form see Lakalai -gi-.

POC *t usually remains /t/.

Examples (in addition to those above): *tok *thud, thump*, S. tok *beat a slit gong*; *toko, S. e-tok *walking stick*; *tutua, S. tut *beat*; *saqit, S. sihit *sew*. The only exceptions noted are S. hisik *sea* (see note 15) and possibly hiŋis *cloud, sky* which may represent a borrowing from a language of the Kimbe group (cf. Bali laŋiti) in which /t/ before /i/ shifts towards /s/ (as in Lakalai).

POC *q is reflected as either /h/ or /k/.

Examples: *quma, S. kum *work*; *saqit, S. sihit *sew*; *muqa, S. muh (for discussion, see 5.3)

As in many languages, reflexes of *l, *d, *R vary, and often not enough examples have been found for a pattern to be established. The most common reflexes are /l/ and /h/.

POC *l is usually S. /l/.

Examples: *laŋo, S. e-laŋ *fly*; *lumut *moss*, S. lumlum *moss on tree*; *keli, S. kel *dig*; *sulu, S. e-sul *torch*, etc. But possibly *l is reflected as /h/ in the second syllable of S. kihon *hear*. See also PAN *mulut *mouth, snout*, S. muhut *nose, snout*.

POC *R produces S. /l/ in *kaRati, S. k(a)lat *bite*, and possibly in *maiRa(q) *red*, S. m(e)lek *flame, glow*, but /s/ in S. misi *urinate* and (Ø) in S. nekoit *octopus*, discussed above.

The cognacy of items reflecting POC *d is uncertain except probably for *dadasi *peel, scrape*, S. las *undo, flay* (and see also S. la| *scratch*), and POC *somod *dirty*, S. sumuh *dirt*.

Vowels are usually close to or identical with POC vowels (allowing for my problems in hearing and recording, mentioned above), apart from the effects of umlauting. Sometimes an expected shift has not occurred, as with *tanum *bury*, S. tanu *cemetery*, suggesting that final consonants had been lost in these cases before the second vowel could affect the first.

Like several other Oceanic languages (including Molima and Kove) /a/ is fronted when immediately following by /i/, and I have assumed that me, mei *come* reflects this process, as possibly does S. m(e)lek *flame, glow* (if it is indeed from *maiRa(q) - but the final vowel is then

assumed to have been affected by the preceding one). See, however, S. *mamai-tongue*.

It must be added, however, that several possible OC forms have not been put forward because they contain unexpected vowels. These include S. *te faeces* (POC *taqe), and S. *malel light in weight* (POC *maRaRa). Sometimes, of course, there is exterior support for an unexpected vowel; S. *tapu grandmother* has an unexpected first vowel that is also found in the Tolai and Kiriwinan words for *grandparent*.

Finally, it should be noted that the special forms *moi come* (2nd p. only) and *lo go* (2nd p. only) may reflect influence from the subject pronoun.

4.2 Discussion

In Sengseng the most obvious tendency is a reduction of many roots to monosyllables, typically with the shape CVC or CCVC. In the majority of cases, this is accomplished by dropping the final vowel from a root of the shape CVCV. Here I am assuming that the root was in the form reconstructed for POC, often ending in a vowel, rather than the PAN form reconstructed for many verbs, in particular (see Blust 1977). To take examples of verbs beginning with /k/, which reflects a similar stop in POC, we find: S. *kak broken, as a book* - POC *kaka; S. *kap pick up, as with tongs*; e-*kap tongs* - POC *kapi; S. *kas scratch, rub, itch* - POC *kasi; S. *kel dig* - POC *keli; S. *kok (or kuk) carry* - POC *koko; S. *kom clasp* - POC *komo; S. *kot cut* - POC *koti.

If the root began with two identical syllables, the first might be dropped. Examples include S. *mak chew betel* (PAN *mamaq), S. *las undo, skin* (?POC *dadasi), and *misi urinate* (POC *mimiR). In the last case, though cognacy is certain, the final vowel casts some doubt on whether the form is directly inherited. See also S. *tut beat, as barkcloth* (POC *tutua) and *pup break wind* (POC *puput), which go against the rule just mentioned but do produce a verb of the most common shape.

Some other examples suggest that in other cases, the first syllable might be dropped even when it was not identical with the second. The case of *lit skin* (?PAN *kulit) has been mentioned above, and other possible examples are S. *suk point to, indicate*; e-*suk index finger* (?POC *tusuk) and S. *kut tail* (?PAN *ikuR).

Sengseng nevertheless contains many disyllabic roots which end in vowels. Parts of the body of course take suffixed possessives, and the final vowel of a root may be dropped or retained depending on the form of the possessive pronoun, so that final (and unpronounceable) consonant clusters are avoided. The verb meaning *suck milk* is *sus*, but the similar word for *breast* retains its final vowel in certain contexts: *sus-it our (inc.) breasts*; *susu-n its breast* (note the shortening of these possessive forms). Verbs, and nouns derived from verbs as in some of the examples given above, preceded by one of the noun-markers (e- or a-), are most likely to be monosyllables. In many cases, however, the final consonant of a PAN form has been retained, but if the shape is CVCV and if the two vowels are identical, the first one may be dropped in ordinary rapid speech producing an initial consonant cluster. Examples include S. *kalat, klat cut through, bite* (POC *kaRati) and perhaps S. *melek, mlek to light, flame* (?POC *maiRaQ).

When the vowels are not similar, a shift may take place in which *a in the first syllable shifts to conform to /i/ or /u/ in the second, but only if the word ends in a consonant. Examples are: S. *sihit sew* (*saqit) (but with a possible doublet *sak tie*); *tinis* (PAN *taŋis);¹⁴ *hisik sea* (POC *tasik);¹⁵ possibly S. *hiŋis sky, cloud*, (POC *laŋit) despite the unexpected final consonant; *humuk gnat, mosquito*, etc. (POC *ŋamuk); *mutuh mature, old* (POC *matuqa). This type of umlauting does not necessarily reduce a word to a monosyllable; such reduction can occur with words like *sihit*, because both /h/ and /v/ between like vowels may be dropped in normal speech, but most of the forms just cited would have unpronounceable initial consonant clusters if the first vowel was dropped.

4.3 Problematical reflexes

It should be evident from what has been cited so far that PAN forms are relatively easy to identify with a fair degree of certainty. The reflexes mentioned, although they may show inconsistencies probably reflecting borrowing between adjacent languages (see below), are not hard to accept. Another postulated set is perhaps less convincing. One of them rests on the supposition that *p in some cases became S. /y/. Examples include S. *yak to fly* (POC *kape, from PAN *pakpak with metathesis?); S. *e-yah* (rare), *axe*, with doublet *pak cut wood* (POC *paqa); S. *yaŋyaŋ disperse* (PAN *paŋpaŋ); and possibly S. *yoŋ together with (another person)* (PAN *punpun). There can be no doubt that the usual reflex of POC *p is a bilabial stop (e.g. *pa- causative prefix*; S. *pan to paint* (POC *pani); S. *put pluck* (POC *putik)), or occasionally a bilabial spirant or semi-vowel. The reciprocal *-val* (POC *paŋi) and *a-val hut* (POC *pale) have already been mentioned; another possibility is *hvok rotten* (POC *popok; cf. *Miu pohok*; *Karore vrok*). These reflexes of *p are common in Oceanic languages but to my knowledge a /y/ reflex is not, and I am somewhat more dubious about the proposed cognates in this case. Johnston (1983) has proposed other S. reflexes for his south-west New Britain proto-phonemes *p and *mp, such as S. /k/ for the former (one example). The words given as examples, even if genuine reflexes of proto-forms, cannot now be attributed to Oceanic, and do not affect my generalisation.

Many proto-phonemes (though, as usual, not *m) produce more than one reflex in Sengseng. As far as I know, this is universally true in Oceanic languages, and however much the alternatives may be attributed to borrowing, they do not usually impede classification of the language as AN (see Chowning 1973:197-200 for examples from Lakalai and Kove) unless the variants can plausibly be derived from nearby AN languages, as is the case with NAN Wasi (Peleata) of central New Britain (see Chowning 1969:20). Before dealing in detail with borrowing, however, I want to mention one peculiarity of Sengseng. PAN *k is almost always reflected by S. /k/; the exceptions are, as in many languages, the pronouns (see Table 1) and one or two cases which are dubious because of the odd reflex, such as *e-mut louse*. By contrast, in a number of cases PAN *g is reflected by S. /ŋ/: S. *ŋi tooth* (PAN *gigi(q));¹⁶ S. *ŋilŋil shiny* (PAN *gilaŋ, *gilap); S. *kuluŋ, kluŋ thunder* (PAN *kuDug); S. *ŋep pant* (PAN *gapay weak); S. *ŋaŋ call, cry out* (PAN *gaŋ bark). I am not seriously suggesting that Sengseng is not an Oceanic language, but the data seem worth putting on record.

Table 1: Sengseng pronouns

Independent and preverbal (except questions)				
	Singular	Dual	Paucal	Plural
1 inc.	ŋa*	taŋa, toŋ	souka, souk	ita
1 exc.		toha	piok	pima*
2	o, a	mom	miok (also numeral "3")	om, ami*
3m.	vi	hiloŋ	hilok	po
3f.	et			po
3n.	i, li			li
Suffixed possessives				
1 inc.	-ŋo			-it
1 exc.				-pim
2	-ŋon, -p(rare), -m(rare)			-m, -om
3m.	(vi) ... -n			-hi, -po
3f.	(et) ... -n			
3n.	(i, li) ... -n			
Separable possessives				
1 inc.	ta -ŋo	(Dual and paucal are rarely used. Forms are the same as independent ones, with any final /-a/ omitted.)		ta-it
1 exc.				ta-pim
2	ta-p			ta-m, ta-om
3m.	ta-vi, vi ta-n			ta-po
3f.	ta-et, et(i)ta-n			
3n.	ta-n			
Goal				
Same as suffixed forms, except:				
2nd p.s. -ŋon (contrast ta-p <i>your</i> , and ta-ŋon <i>to you</i>).				
3rd p.s.n. -i, -li, -ni (last only suffixed to a verb ending in /i/). <i>Eat it</i> (vb.i):i-i, i-li, i-ni.				
If a verb has a 3rd p. personal subject and object and they are different, -sun is substituted for the object pronoun: vi hion-sun, <i>he saw the other man</i> .				
Vocative				
2nd p.s.			tatoŋ	
2nd p. dual			hiloŋ	
2nd p. paucal			hilok	
2nd p.pl.m. (and mixed sexes)			apo	
2nd p.pl.f.			epo	
Other special forms				
1st p.s. intensive and interrogative				ŋo
2nd p.s. interrogative				ŋon
1st p. dual interrogative				taŋ

* For discussion of the distribution of cognate forms in AN languages see Lincoln 1978:940,945.

4.4 Borrowing

As will be explained, the pattern of Sengseng word tabus ensures that a considerable amount of borrowing will take place, and it seems probable that the same situation obtained in the past. The degree to which Pidgin words had entered the language, even at an early stage of contact, also suggests that the Sengseng are exceptionally ready to adopt foreign words and phrases even when they are not replacing those that are tabu (in contrast with, for example, the Lakalai, who were very much slower to accept a substantial amount of Pidgin). The question is then bound to arise of whether the AN component in Sengseng can be wholly attributed to borrowing from neighbouring AN languages. As I pointed out earlier, Sengseng itself is in direct contact with only one other non-Pasmanua language, a local dialect of Arawe (referred to by the Sengseng as the 'island' language). Trade and intermarriage, however, ensure a good deal of boundary crossing, and other Pasmanua languages lie adjacent to those of Lamogai, to the west, and approach the Kimbe languages to the north.

The first point to make is that none of the languages of south-west New Britain, except when directly adjacent to one of the exemplary languages of the north coast, has (on the surface) a higher AN content in the basic vocabulary than does Sengseng. So the AN content of Banaule (Kapore, Bebeli) is relatively high because of the many loan-words from Lakalai (hence Dyen's suggestion that they might belong to the same subgroup - 1965:47-48) and Aria, of the Lamogai languages, shows the same effect because of loans from Kaliai. Both Arawe and Kaulong-Sengseng contain a number of AN forms in the basic vocabulary, but if we consider dialects of Arawe that are not adjacent to Kaulong or Sengseng, the only probable AN forms that have the same shape in both languages are *sinaŋ swi*, *day*, *a-mat snake* and *me come*. The first two of these are found throughout south-west New Britain, and form part of the evidence that led both me and Johnston to suggest genetic relations between the Lamogai, Arawe, and Whiteman languages (see below). (I am omitting forms found only in 'married' speech in the Pasmanua; see below.) Other AN forms in Sengseng vocabulary either are not recorded for Arawe or have such a different shape that borrowing is unlikely. Examples include S. *num*, Arawe in *drink*; S. *tinis*, Arawe *teŋ weep*; and even S. *ita*, Arawe *ta we inc*. In at least one case the Sengseng word is clearly AN while the various Arawe ones are not: S. *hisik sea* beside Arawe *malaŋlo*, *pepek*, *urvu* (from different word-lists). In the case of the word for *sew*, the shape of the Arawe form suggests that it was borrowed from another Pasmanua language, possibly Karore (Arawe *siririne*, Karore *sirit*), whereas the Sengseng form is, as was shown above, regularly derived from **saqit*.

4.5 Direct inheritance or borrowing: Johnston's PSW

In 1983 Johnston expanded on his own 1982 paper and the 1981 paper by Lynch to suggest that all the languages of south-west New Britain which I had originally (1969) assigned to the Lamogai, Arawe, and Whiteman Families, as well as a few others such as Uvol (see above) derived from a common ancestor which he called Proto-Southwest New Britain (PSW). As I noted above, he includes Uvol in a subgroup (Western Whiteman) with Sengseng and the other Pasmanua languages, an assignment which I reject. Johnston proposed a number of 32 "tentative proto-etyma" for PSW which he derives from POC. Some of these derivations are certainly valid (e.g. PSW **puri banana*, **mata eye*, **monuk bird*) but others seem more dubious (e.g. PSW **kowozak rat* beside POC **kunsupe*; PSW *(z)umpa *big* beside POC **la(m)pas*, PSW **elik fish* beside POC **ikan*).

In addition, numerals and pronouns are assumed to have undergone very peculiar transformations (see below). Of the basic 32 proto-etyms, Sengseng has certain reflexes of seven, one additional in the 'married' vocabulary (see below), and three other possibles, all of which have unexpected shapes: mut *louse* (POC *kutu), kliŋa *ear* (POC *taliŋa), and umat *stone* (POC *patu). Johnston attributes to PSW and ultimately to POC all of these, as well as several other Sengseng words. One of these (kunam *turtle*, derived from POC *poŋu via PSW *(p,k)onom) is not on my list, which has only miyat *turtle*. The others are terms which seem unlikely to be of POC origin: S. kive *leg*, which Johnston derives from PSW *kawē and POC *waqē; S. vo *mouth*, POC *awana; S. homan *new* derived from PSW *taŋan and POC *taŋan (source unexplained); S. hot *neck* [not *throat*, as in Johnston's translation] derived from POC *qolo; and S. miyaŋa *forehead*, derived from POC *ndamwa via PSW *zamwa. In one case, he drops the first syllable from the Pasismanua word for *hand* (S. vili; see Throop and Throop 1980:237), reducing it to li, without explaining what he has done, and then derives it from POC *lima. One or two of these S. words may indeed be POC in origin, but the enormous irregularities involved seem to rule out most of the others, except as possible borrowings. As will be seen, although I think many Sengseng forms have indeed been borrowed from other languages, the burden of the evidence, as presented above, does not suggest very complicated sound shifts for forms which I assume are directly inherited. For further discussion of Johnston's whole theory, see below.

If Johnston is correct, it remains unlikely that Sengseng acquired most of its AN component by borrowing. The distribution of the languages makes it probable that the 'exemplary' ones of the north coast rather than the aberrant ones of the interior and south coast are the late-comers to New Britain (see Chowning 1976:379-380). Lapita pottery in the Talasea region of the north coast suggests an early AN occupation there, and it may be the present largely interior languages (Lamogai and Whiteman) that represent that occupation rather than the present north coast languages. Blust has suggested (personal communication) that the AN content of Sengseng perhaps derived from a "source language (which was there (but) has since died out or changed location". While admitting this as a possibility, the fact that Sengseng and its neighbours seem so like other Oceanic languages in grammar makes me feel that it is easier to assume that at least some of the content that is not obviously AN came from NAN languages which still exist in pockets throughout New Britain and which are generally believed to have reached the island before the AN ones; Johnston has in fact attributed many lexical peculiarities of his SWNB to borrowings from NAN. The low AN content of these languages would then be explained in terms of the internal changes to be discussed below.

It should be noted that there are cases in which other AN languages of New Britain have presumably borrowed from a Whiteman language. For the present, I will stick to Lakalai, flanked by Whiteman languages, which both Goodenough and I have suggested were spoken on the north coast before the ancestors of the Lakalai arrived there. Among other lexical items shared by Lakalai and Sengseng (and presumably by the intervening languages) are names of several wild trees: Lakalai la-uele, S. e-vel *Canarium almond*; Lakalai la-koi, S. e-koi *wild areca palm*; Lakalai la-salumu, S. sa-nuhum *Ficus*; Lakalai la-ropa, S. e-lop *Pometia pinnata*; and of at least one wild marsupial (Lakalai e-misiki, S. e-smik).¹⁷ In most cases it is impossible to ascertain the direction of the borrowing, but the word for *Ficus* in Sengseng literally means *tree-big* (each Sengseng village was traditionally built around a large strangler fig). In normal pronunciation these are Sengseng sanum, Lakalai salum (Lakalai being a dialect in which /n/

in other Nakanai dialects has become /l/), and it seems clear that a language in which the term is meaningful was the source.

The additional lexical items specifically shared between these languages are of three sorts. Some are common words, not known to be AN, which simply occur in both (e.g., S. *ko*, Lakalai *koko* *defecate*) in which it is impossible to suggest either the source (which of course may be still another language) or the age of the item. (M. Ross has recently suggested, though not in print, that this word derives from POC.) Other words, because of their shape and their existence beside more obviously Sengseng forms, are almost certainly recent borrowings, presumably by Sengseng who have worked on the plantations near Talasea. These include *mahela* *shame* (identical in Lakalai) beside S. *manjin*, possibly formed from the verb *man* *to hurt* (from POC **manuka*?) with a noun-formative suffix, and *sesele* *truly* (identical in Lakalai) beside various complex Sengseng phrases involving oaths. The third category consists of adverbs, without synonyms in either language, which seem unlikely to be borrowed easily. These include S. *lai*, L. *lalai* *tentatively*; S. *so*, L. *sou* *yet, still*; S. *akai* (pronounced *agai*) *now, at once*; L. *gai* *in the very near future*.¹⁸ At the very least, such shared forms do point to a period of contact between the close relatives of these languages, but note that the AN content of Sengseng is not much altered by such evidence. Perhaps because Tolai (Kuanua), though not an 'exemplary' language, has been much used for the reconstruction of POC, it is worth making the same point about forms which Sengseng shares with it. Tolai is physically very far from the Whiteman languages,¹⁹ with various NAN languages intervening, but shares with Sengseng several forms that are not normally attributed to POC. These include *kul* *buy* (the same in both languages); S. *kila*, Tol. *gila* *ignorant* (the opposite meaning from POC **kila*); S. *iya-*, Tol. *ia-* *name*; S. *molo*, Tol. *moro* *yellow*; and see also Tol. *kuru* *penis, tail* beside S. *kut* *tail*. Assuming that both languages did not borrow the same forms from intervening NAN languages, they may share some vocabulary that belongs to an early period of the settlement of AN-speakers in New Britain and New Ireland.²⁰

5. WORD TABUS

I mentioned earlier that Sengseng contains a large number of synonyms, and indicated that some of these may be the result of name tabus, which require substitutes. In many cases the Sengseng and Kaulong speakers are quite aware of the source of borrowings, and in other cases it is possible, with some certainty, to trace the source. Deliberate borrowings and a variety of other ways of replacing vocabulary are all part of the same process, and at this point its operation will be described before the source of some of the substitute terms is dealt with in more detail.

5.1 Name tabus

The effects of tabus on personal names for a language in general, rather than simply on individual usage, vary with a number of factors. The most obvious is the degree of resemblance between personal names and other lexical items. Even an extensive set of name tabus will not affect the language if, as in Lakalai, few personal names resemble other words (and if nicknames, which of course do, are not subject to tabu).²¹ (For this reason we cannot

directly correlate the existence of name tabus with effects on the lexicon; cf. Simons 1982:183).

In Sengseng, most people have a special individual name; it is rare, though not unknown, to find two people with the same name, or to name one person for another. What I will call the personal name is usually bestowed at birth, though often changed during adolescence, in which case the earlier name is abandoned. The most common personal names consist of one or two syllables usually followed by *-li*, for males, and *-me*, for females. The tabus do not apply to these common last syllables, which as independent verbs mean respectively *go* and *come*. Names need not have these endings. In addition, many people have nicknames denoting personal characteristics such as a runny nose or prematurely white hair; these inevitably contain everyday words. Men traditionally had a special nickname, which was supposed to be kept secret for women, which referred to the appearance of the man's penis, likening it to a plant or animal. Occasionally such a name comes into general use, as with one prominent man called Kaŋkaŋ (Hornbill) — and perhaps another called Wul (Eel) — in which case women were not supposed to know the origin of the name.²² In addition, when boys have their teeth blackened each is given a new name which ends with *-kit*, the word for tooth-blackening material. If syllables of this name are tabued, the word *kit* is not. The only names not subject to tabus are baptismal names assumed by converts to Catholicism, which now embraces most of the people living near the coast.

For both sexes, it is tabu to say the name of any affine of senior generation, or any word resembling that name or one of its syllables (other than *li*, *me*, *kit*). The consequences of breach of the tabu are, in theory at least, disastrous: sickness, possibly the death of the spouse, and failure to have children and to gain wealth in shells, the two great desires of individuals of both sexes. The tabu extends across languages; a person who cannot say *e-ki water*, *bird*, also cannot say Pidgin *ki key*. Goodale was told in Kaulong that the tabu lasted until the death of the affine, but she may have misunderstood; in Sengseng it lasts until the death of the person observing the tabu, regardless of whether the affine is alive or not.²³ It applies not only to living affines, but to well-remembered dead ones, such as those buried within the hamlet's men's house or identified as the planters of fruit trees around the hamlet. I knew three women, married to a set of brothers (though one brother was long dead), who had to avoid *homan new, now* because it was the name of their husbands' grandfather, who would certainly not have been alive when they married. Normally the residents of a hamlet centring on a men's house are cognatic kin and so likely to share at least one remembered ancestor, so that all in-marrying women may have some name tabus in common. In addition, if a tabu name includes the name of a food, it is also tabu to eat the food. In time it is possible to lift the eating tabu, like some other affinal tabus such as approaching the grave of a dead tabu affine, by giving a pearl shell to the spouse's kin, but the name tabus cannot be lifted.

In consequence of the name tabus, the Sengseng (and Kaulong²⁴) languages contain lexical items called 'married talk', substitutes for many common words. The people actually describe the situation as if there are two distinct languages separating the single from the married,²⁵ and as if all the married use 'married talk'. In fact, a married person need not use the substitute form unless it appears in the name of a tabu affine, though many do so as a matter of course. (The test is whether such a person will say the 'single' or 'true' word if necessary, as in speaking to someone like me who may not recognise the alternative.) When first married, people may simply shift

vocabulary to avoid giving offence until they learn who all their tabu affines are. An extreme form of shifting vocabulary was practised by a newly married Kaulong man, more widely travelled than most, who told Goodale that he tried to avoid speaking Kaulong at all, substituting Pidgin and the language of Talasea where he had been a labourer (and thereby rendering himself largely unintelligible to his bride).

At present, in this area, name avoidance cannot be accomplished by minor alterations in pronunciation, as is the case in Kove (where, for example, they substitute voiced stops for spirants). Certain sound shifts between neighbouring languages are well known, such as Karore /r/ for Sengseng /h/ in many words, but a Sengseng avoiding the tabu word *sew* could not just substitute the Karore version *sirit* for *sihit*. As far as I know, metathesised forms, which are frequently heard (e.g. *ahu* alongside *hau cry out*) do not serve to evade tabus, but I failed to ask about this. For many words, a number of standard married forms exist and are generally known, so that in Dulago village the married forms for *pig* (*yu*) include *a-ni*, *kanem*, *a-tiem*, and *pasi*. This set is interesting because the origin of the words can be traced, though the Dulago people did not do so. In Angelek village in Kaulong, *a-ni* is a modifier used to distinguish wild from domestic pigs, and *kanem* is known to be a borrowing from Lamogai languages to the east. *A-tiem*, which is said in Sengseng also to be a word for *ghost*, is used to talk about pigs when they are being carried through spirit places so that the spirits will not be attracted to the pig. Throughout West New Britain, cognates of *pasi* are general terms designating a number of game animals, including pig in Kove *basi* but in Lakalai referring only to smaller animals such as possum.

Each small region in the Pasismanua has its stock of married terms, which vary locally. For example, I found that although people in the 'coastal' Sengseng village of Suvulo recognised most of the married terms that I had learned in the interior, some were quite unfamiliar to them. Goodale's impression was that much of the perception of dialect differences in Kaulong derives from differing sets of married forms. These standard substitutes do not cover all the possibilities, however, and often a newly married person must improvise. Failing knowledge of a sufficiently different term from a foreign language, he may use other methods. One is a slight shift in meaning or narrowing of range, as in the pig terms just mentioned. Another is a greater shift that is still readily intelligible, such as *tongs* for *hand*, and *vine* for *snake*, or the complex construction (time marker + completion marker + *here*) that the three Dulago wives used for *now*. The next two groups seem to be probabilities. The list of Sengseng synonyms contains many nouns formed from verbs, as *kesjin cutter* for *knife*, *nasjin the chewed* for *sugarcane*, and *pahihjin shaver* for *obsidian*, all of which may have developed as someone's married substitute for the usual term. In addition, many lengthier descriptive terms exist for animals and plants, again as alternatives to shorter terms. For example, one of the numerous words for wallaby means *soft fur*, one for python means *crooked middle*, and one for candlenut means *spear polisher*. Again, it seems likely that these alternatives to shorter unanalysable terms were originally developed in response to name tabus; certainly I have no evidence of other origins (as in poetry or magical spells). Finally, in some cases a person seems simply to invent a word, as personal names and most tooth-blackening names are invented, and lets others learn its meaning from context.

It should be noted here that along with this proliferation of synonyms, the Sengseng fail to make many distinctions which I am familiar with in other languages. The word translated 'spear' above also embraces 'wood', 'tree',

and a variety of wooden tools such as the trigger of a trap, and the synonyms usually have the same range. A single word for tree bark is used for a bark cooking roll, a bark umbrella, and a bark baler for a dammed-up stream. Presumably this absence of many specialised terms prevents the number of words each individual must learn from reaching an unmanageable size. As it is, the unmarried especially enjoy displaying knowledge by listing all the synonyms they know.

Contrary to what one might expect, in view of the fact that children first learn married forms from their parents, the Sengseng consider the 'single' form to be the basic original word for something.²⁶ Often when I was being told a place name by a person who had to use a married form for one of its syllables, I was urged to supply the single form myself so that I would record the true name. It should be added here that many Sengseng men never marry, and others delay marriage until they are at least middle-aged (see Chowning 1980), so that some men never use married terms after early childhood, and others use single terms for most of their lives. Traditionally, however, all women are married, and usually at a fairly young age, so they are perhaps to be considered the main sources of linguistic innovation and change. Given a common though not invariable post-marital pattern of virilocal residence, women are also more likely to have to avoid the names of dead affines buried near where they live. On the other hand, most marriages take place between individuals who not only live near each other but who share at least one common ancestor, so men also often interact with affines.²⁷ The evidence indicates that in some cases, the married term completely replaces the original, and I suspect that many such cases occurred locally prior to pacification when incessant warfare limited contact between small local groups.

5.2 Other word tabus

In addition to the affinal name tabus, there are others that affect vocabulary. At least in theory, all Sengseng within the vicinity tabu the name of anyone who dies for a period variously said to range from a few days until decay of the flesh. I frequently heard breaches of this tabu; if it once was influential, it no longer seemed to be. (In many parts of Melanesia, tabus on names of the dead are more influential than affinal tabus because they often are observed by all members of the community; see below and Simons 1982.) Some tabus are even more temporary, but do add to the local store of synonyms. In certain places inhabited by spirits, particular terms are avoided that might anger or attract them, and sometimes a substitute is used, as with the a-tiem pig term mentioned above. When collecting shrimp, it is necessary to avoid their usual name (e-lus) lest they hear and flee, and the name of a red leaf is substituted. In general, spirits tend to be addressed and sometimes referred to by kinship terms, notably tisa *grandfather*, and this has become a married term for fire, which is both personified and sacred. Finally, avoidance of sexual terms and other obscenities in mixed company has led to the use of various euphemisms. Men say that women have a secret vocabulary for discussing sexual matters, but I did not confirm its existence. Conversely, though a man could be killed for using a Sengseng obscenity in the presence of women, many delighted in using what they thought or said were obscene exclamations borrowed from other languages (on the assumption that the women, who travel less often, would not recognise them.)

5.3 The acceptance of substitutes

For tabus, including euphemisms, to become fixed in the language and actually replace earlier terms, they must of course be known and used by most people. The small largely inbred populations in south-west New Britain present an excellent opportunity for replacement, as what was originally a substitute term comes to be so widely used that it is thought of as the original. That this has happened is shown by the fact that in some Kaulong villages, including Angelek, e-mon, elsewhere the prime example of a married form for bird, has become the term used by everyone. Goodale reports other cases in Angelek of common use of what were substitute terms in the interior village of Umbi, though the affinal tabus were still observed in Angelek. In Suvulo I found that in several cases a Pidgin word such as kambang *lime* had not only become the usual term but was firmly stated to be Sengseng in origin. Of course such cases of replacement by borrowing occur where no name tabus exist, as in Kove, but the opportunities do seem considerably greater in a situation like Sengseng, in which substitute forms are constantly sought, created, and widely used.

In eastern Oceania, particularly, word substitution as the result of name tabus can easily become established when the names are those of chiefs, so that everyone has to observe the tabu. Widespread tabus, as I noted above, are more likely where the names of the dead rather than affines are avoided, even when the dead are not chiefs. In parts of the Massim, where the kin of the dead all take offence if the name is spoken, such tabus can indeed alter the language, as Lithgow (1973:106) records for Woodlark. Where supernatural sanctions back up the tabu on names of the dead, the effect can be even stronger with people only peripherally related, who do not know the exact connections and avoid the name out of fear of offending, as Keesing found in Malaita. The Kwaio tabus in any case affected all members of a cognatic descent group and locality, and could spread by out-marriage and fissioning, producing a situation in which "there are significant variations in vocabulary (between localities) many of which resulted from word tabooing" (Keesing and Fifi'i 1969:171). Here is certainly one possible explanation for the amount of linguistic diversity in Melanesia, which has long seemed difficult to understand. Although I agree with Pawley (1981:273-275) that a major reason is simply normal divergence over a long period of time, he does not really tackle the problem of extreme diversity in a geographically compact area like New Britain except by postulating "local movements and the intrusion of alien, enclave communities" (p.275; compare the statement on p.289 that "major subgroups tend to correspond fairly well to a discrete island group"). Here I would like to suggest that the great diversity found in regions in which the continuous distribution of closely related languages argues against migration as an explanation, may owe something to the widespread Melanesian phenomenon of incessant warfare between communities with a single language and culture (see Chowning 1977:41-42). This both weakened the likelihood of political confederation and increased the possibility that small local differences would become fixed. Name-tabuing could then have a chance to lead to linguistic diversification even in the absence of chieftainship. I am not suggesting that the situation just described was found only among AN-speakers; on the contrary, NAN-speakers may not only have undergone linguistic change for the same reason (see Franklin 1977:13-14), but western Melanesian political behaviour may owe something to influence from people resident in the region before the AN languages arrived. If Pawley is correct (1981:285) in believing that hereditary leadership can be attributed to a Proto-Melanesian stage, its rarity in the west needs explaining; contact with speakers of NAN languages may have contributed both to the development of the

Big Man system so typical of the west, and to their patterns of warfare, both of which may have helped accelerate linguistic change throughout this area.

On the other hand, the same small size that facilitates linguistic differentiation can work against differentiation, because members of small populations must often seek mates outside (even where marriages do not result from sexual attraction between relative strangers met at large gatherings, as they sometimes do in Sengseng). Intermarriage, which can also lead to adoption across linguistic boundaries,²⁸ must be a major reason for the existence of differing pronunciations of the same word, and possibly of the great variety not only of Sengseng kinship terms, but of possessive endings for them. Often the source of the variation can be identified. For example, in many words Sengseng /-h/ corresponds to Kaulong /-k/, as in S. *muh*, K. *muk precede* (POC *muqa). Consequently when faced with such alternate forms as e-sih, e-sik *snare* and e-yah, a-yak *axe* it is probably safe to assume that the latter was originally a Kaulong pronunciation, even if it is no longer so identified by the Sengseng (only the forms with -k appear in the Kaulong word lists that I have seen). Possibly when one pronunciation, not necessarily the original one, became preferred, we have the explanation for varying reflexes of a single proto-phoneme (see 4.1), as S. *mak chew betel* where we might expect *mah*. Again, with kinship terms it is possible to identify *wohuk my brother* as Kaulong because both the first syllable and the ending are characteristic of that language. Often, however, one can only postulate an outside origin without identifying it. To indicate the diversity: Sengseng has four different first person singular possessive endings (address and reference) with kinship terms (-a, -h, -k, -ŋo, as well as *veheŋ mother's brother* which either has \emptyset or a fifth ending), and there are three for second person singular reference (-m, -ŋon, and -p), the last being the usual Kaulong (and -ŋon the usual Sengseng) form for suffixed possessives (see Table 1). The Sengseng words for *father* and *father-in-law* take -p, for *mother* and *brother* -m (of uncertain origin, although of course AN), and for *sister* -ŋo. The reasons for this proliferation elude me, but out-marriage seems the most likely explanation.

What emerges in the Pasismanua languages, including Sengseng, is the situation that Grace has recently (1981:266) described for New Caledonia, with "complicated sound correspondences ... due to large scale borrowing of core vocabulary from related languages". In this paper I am arguing that irregular correspondences should not prevent us from accepting certain forms as AN, since we have evidence of extensive borrowing between related languages in this region. More importantly, however, I suggest that the present pattern of word substitution because of name tabus, coupled with the large number of synonyms in each language, indicates that much vocabulary has been replaced – and in some cases undoubtedly by AN forms from other languages. Nevertheless, most of the present AN content of Sengseng and its neighbours cannot plausibly be derived from other AN languages now present in New Britain. In view of the grammatical evidence, these seem to me most likely to be AN languages that have simply undergone rapid lexical change of a type which has probably been common in the small societies of western Melanesia, but which, as Keesing and Fifi'i, and Simons, pointed out, can cause major problems of interpretation when lexicostatistical data are relied upon.

The conclusion that follows is that the number of AN lexemes retained in the 'basic' vocabulary²⁹ cannot be used to determine with certainty whether a specific language is to be classed as AN. In his 1981 paper, Blust points out that he does not know why retention rates vary so greatly between unquestionably AN languages. If, for his sample, the rate ranges from 59% for Malay to 16% for

Gapapaiwa of Milne Bay, there is no theoretical reason why it should not fall well below 16% for languages that are still AN, according to other criteria. A considerable AN content in the remaining vocabulary and in grammar, particularly when this cannot plausibly be derived from other AN languages in the vicinity, should outweigh the single criterion of the content of a word-list. It follows, of course, that in some cases AN languages may have fewer words of AN derivation in such a word-list than some NAN languages, as with the famous case of Mailu: for West New Britain, a possible example is Wasi, originally considered AN by several observers (see Chowning 1976a:189, 190).

6. THE PROBLEM OF JOHNSTON'S PROTO-SOUTHWEST NEW BRITAIN

I have argued that the combination of grammatical evidence and a substantial amount of AN lexicon, not attributable to borrowing from the unquestionably AN languages of the north coast of New Britain, make it plausible to call the Pasismanua languages AN. On the evidence available to me, the same argument applies to all the other languages of south-west New Britain: those that I have called Arawe, Lamogai, and Whiteman, and that Johnston has called Arove, Lamogai, Eastern Whiteman, and Western Whiteman.³⁰ They look neither more nor less AN than each other, except where they have been influenced by adjacent north coast languages. The questions then remain of whether they all derive from a single relatively recent ancestor, Johnston's PSW, and whether specific phonological changes from this putative ancestor (rather than from POC or PAN) justify Johnston's subgroups.

To begin with, two points must be repeated. First, I reject the inclusion of Uvol in the same subgroup with Sengseng, and so shall be discussing only the remaining Western Whiteman language, my Pasismanua. Second, I accept that certain isoglosses extend through the languages of south-west New Britain: e.g. forms like S. amat *snake*, sinaŋ *sun*, miok *three* and nal (not ŋal, as in Johnston) *four*. The first two of these seem to me unquestionably AN, but a-mat can be derived without difficulty from POC *ŋmata, either with metathesis or with a prefixed article, and if Johnston is correct in postulating a POC form *nsinaŋ *sun*, the same argument applies to it.³¹ The second two do not seem to me OC, despite Johnston's ingenious derivation of PSW *moiook from "metathesised forms apparently reflecting POC *kamiu 2 pl. in all SWNB" (Johnston 1983). The Sengseng trial/paucal pronouns (see Table 2) all have an -ok ending except for souka *first person inclusive trial*, and miok *is*, as well as the word for *three*, also the second trial form. (The -ok ending also appears in the pronoun yok designating three people together, as opposed to yoŋ for *two*; for -oŋ as a dual marker see Table 1.) It is, of course, commonplace for dual and trial pronominal forms to incorporate the words for *two* and *three*, but if we derive S. miok from *kamiu, we are left with the second plural pronouns om/ami which can be more easily be derived from *kamiu without invoking metathesis.

As regards the Pasismanua languages, the forms postulated by Johnston seem to me to fall into four categories. First are those in which PSW is identical with POC, such as *mata *eye*, *tama *father*, *taliŋa *ear*. Second are those with no recorded reflexes in the Pasismanua languages (excluding Uvol), such as PSW *puri *banana*, *towu *sugar*, *izuŋ *nose*. Third are those for which Johnston has identified Pasismanua reflexes which are unacceptable because of incorrect data, including mistranslations and false divisions of words, and because of unlikely sound shifts. As well as the word for *night* mentioned above, examples include confusion of words for *forehead* (S. miyaŋa) and *head*

and the derivation of both from *zamwa *forehead* (POC *ndamwa), and the derivation of S. iya-n *its name*, written by Johnston yan, from *azaŋ. So far I have been discussing PSW forms he derives from POC, but the same objections apply to some proto-etyma postulated for PSW alone. I am unpersuaded that some of the forms he derives from a single proto-form are all cognate with each other: for example, that S. tahe-n (his tahe-n) *one* and Arawe *ke* both derive from PSW *kai, or that S. yut *hair* and kin *leaf* both derive from *kañin. Finally, Johnston has proposed a few proto-forms that do seem to be distinctive, whether or not they derive from POC, and to have reflexes in different subgroups of south-west New Britain languages. These include *ñumuk *mosquito* (for POC *ñamuk), *pa-ŋal *four*, and *zeki *water*. In most cases, not enough comparative data are available to make it certain that the supposed reflexes are just that — for example, that S. e-ki *water* and Arawe rei *rain* both derive from the same root.³² In others, the distribution of the distinctive form outside this region is not adequately described; for example, umlauting of the type that would produce *ñumuk is found not only in Kilenge-Maleu but in at least one language of the Siassi Islands (see Chowning 1976:371).³³ Furthermore, we are not yet in a position to assess the importance of borrowing in affecting the distribution of these forms. Several that do not appear in Johnston's list for Pasismanua languages are in fact found in Sengseng either in the 'married' lexicon (e.g. kwon *fire* beside 'single' yau) or as one of the synonym sets (e.g. hvo *two* along with wuoŋ and ponwal).

On the basis of the evidence presented, I cannot see that Johnston has yet proved that all these South-West New Britain languages show evidence of phonological changes that justify grouping them together, possibly with a putative Coral Sea Cluster (Lynch 1982), and separating them from other New Britain languages. Of the five possible phonological and morphological innovations said to "identify PSW as a distinct proto-language and the SWNB group as a group with common inheritance" (Johnston 1983), the Pasismanua languages completely lack three: reflexes of PSW *tilu *three* (with vowel change from POC); a "3 ps alienable possessive form *ka-∅ or *ka-a"; and "first person plural inclusive possessive forms *ka-ri, *li-ri, and *-ri" (see Table 1). Johnston acknowledges that the fourth "innovation", "the accretion -ŋal in PSW *pa-ŋal 4" is found in Bariai and Vitiaz Straits languages, and can only suggest that these latter borrowed from SWNB, but in any case the Pasismanua languages lack the pa- prefix. Johnston does, however, consider "metathesis of the vowel sequences POC *kamiu 2pɿ > kamui > komui > PSW *omu ... definitive of PSW". See, however, the Sengseng forms in Table 2, and also, for at least part of the proposed unique metathesis, Molima (Milne Bay) omi'a 2 pɿ.

All this is not to deny the possibility that the Pasismanua languages do indeed subgroup with others in south-west New Britain (apart from those I originally assigned to the Whiteman Family). In order to prove such connections, however, we need much more evidence in order to establish ties above the level of dialect chains. Even for these, extensive borrowing obscures the picture (and doubtless accounts for much of the disagreement between Johnston and me).

7. CONCLUSION

Without further investigation, some major difficulties encountered by those who have tried to classify the languages of south-west New Britain cannot be ascribed to the operation of name tabus over a long period of time. We cannot know precisely how long these people have been observing their present system of word tabus or handling them in precisely the same way: indeed, I am suggesting that they may once have practised avoidance by the use of minor phonological changes, a method that is not acceptable nowadays. Furthermore, we do not yet know how name and other word tabus operate in other languages of south-west New Britain, though in view of various cultural similarities among these societies (see Chowning 1978a), it is likely that they are similar in these respects as well. Certainly name tabus are not the only factor to have affected lexicon in this region. Some of the diversity must have been caused by intermarriage between small populations, and quite possibly there existed in the past the influence of speakers of NAN languages, now gone from the immediate region.³⁴ Allowing that all the local peculiarities did not have a single cause, if we accept that much of the AN content of these languages is directly inherited, the situation found in the Pasismanua languages at least is strikingly close to that described by Keesing and Fifi'i and Simons. The irregular sound correspondences, the proliferation of synonyms and doublets, and the small number of AN forms in the lexicon coupled with grammar that is by no means aberrant for OC, all fit Simons' predictions about the effects of lexical change resulting from word tabu (Simons 1982:189-190). Simons warns against the danger of offering word tabu "as a panacea to cure all comparative problems" (1982:191), and undoubtedly he is right to do so. Nevertheless, in view of the demonstrable effects of such tabus in the recent history of the Pasismanua, it seems highly likely that these languages look aberrant primarily because they have undergone rapid lexical change, and that a major reason for this change was the operation of word tabus in small scattered populations. Like Keesing, Fifi'i, and Simons, I should expect that similar accelerated change, with the same causes, has affected many languages of Oceania.

NOTES

1. Loukotka considered these languages, along with the Arawe ones, Papuan, whereas Capell called both groups "semi-AN". I discussed their classifications, and rejected them, in Chowning 1969:23. More recently, Blust has informed me that because of the small number of reflexes of PMP forms in basic vocabulary, he considers Kaulong and Sengsen at best "indeterminate" — that is, not proved to be AN.
2. A is a noun-marking prefix which precedes all masculine proper names and many place names, as well as many ordinary nouns. Because it is clearly viewed as a separate prefix, I decided to discard it in writing the name of the language, but I have not heard it omitted when Sengsen and Kaulong speakers are referring to the language.

3. Presumably because they relied on informants unfamiliar with the region east of the Andru, the Throops erroneously identified Karore with Palik, and indicate the Karore region as occupied only by Sengseng (1980:236). Word-lists make it clear that their Palik is an Arawe language spoken on the small islands (other than Kaveng) between the mouths of the Andru and Johanna Rivers. In the recent past, Karore was spoken in the region near the coast around the Ursula River, and (as it still is) on Kaveng Island, just offshore. With the recent moves of some interior Sengseng villages nearer the coast, some formerly Karore-speaking regions are now occupied by Sengseng speakers, as in Suvulo village, and other Karore villages, such as those officially called Pariwa and Mai-ieo, now contain many Sengseng speakers (and are listed by the Throops simply as "Asengseng").
4. I have briefly touched on the subject to be discussed here in earlier papers, initially in one entitled "The languages of south-west New Britain", delivered to the 11th Pacific Science Congress in Tokyo in 1966, and later in Chowning 1976:372. My own fieldwork in Sengseng was financed by the Columbia University Council for Research in the Social Sciences (1962), the National Science Foundation (1963-64), the Australian National University (1966), and the Internal Research Fund of Victoria University of Wellington (1980-81).
5. An outsider might also have difficulty recognising Pidgin words, both because of local pronunciations and local perceptions of word boundaries, which led the Sengseng to regard the initial syllable of many Pidgin nouns as a disposable article, producing such forms as *les* from Pidgin *wailles wireless*, and *morol* from Pidgin *botoi bottle*.
6. The shared cognacy rate between a Sengseng village which contained no foreigners (Dulago) and a Kaulong border village which contained several Sengseng speakers (Umbi) was 80%. Away from the border area, speakers of both languages said that they could not easily understand the other. Johnston has recently (1983) proposed Uvol as another member of his "Western Whiteman", otherwise composed of Miu, Kaulong, Sengseng, and Psohoh (represented by the Bao dialect). Although it was resemblances between Uvol and Banaule (Bebeli), which Johnston now assigns to Eastern Whiteman, that made me originally (Chowning 1969:32-33) suggest a possible link between the Mengen and Whiteman families, I see no reason to accept the assignment of Uvol to the same subgroup as Sengseng. The interested reader can consult comparative word-lists in Johnston 1980b:220-221 and Johnston 1983.
7. I have here rendered Sengseng consonants in my phonemicisation, which uses /p/, /t/, and /k/ for stops which are voiced and in the case of /t/ often trilled when preceding a vowel (in the 'beach' dialect, trilled in other positions as well), while my /v/ represents a bilabial fricative before /e/, /i/, and consonants, but in the interior dialect /w/ before /o/ and /u/, varying before /a/. In the beach dialect the fricative may appear before /u/. Kaulong forms are taken from those recorded by Goodale and the Throops, and not phonemicised. The comparative lists published by the Throops (1980:257-259) relies for "Asengseng" on one that I supplied; pronunciations of Kaulong and Sengseng do not differ as much as the spellings suggest.

8. The term also means *face*, as in many other OC languages, and a homonym means *point (as of a spear)*. See also *snake*. The hyphen following a noun shows that it takes a suffixed possessive. Some S. nouns, but not all, are preceded by articles *a-* or *e-*.
9. I was dubious about this one because of the initial consonant, but Blust tells me that such a shift for **t* before /*l*/ is not uncommon in AN languages.
10. Interestingly, Molima and Sengseng also share the device of an intercalated /*n*/ between a verb root ending in /-*i*/ and a third person singular objective pronoun *i*: S. *i-ni eat it*; *vi-ni hit it*; Mol. *ai-ni-ya ate it*, *vai-ni-ya married him/her*. Because this /*n*/ appears only in this context, I have not interpreted it either as the reappearance of an original final consonant or as a transitive marker (see Pawley 1973:128ff.).
11. cf. Bola of the north-coast Kimbe languages, with first person *ga*, second person *go*, third person *ge* future markers – Johnston 1980b:119.
12. I also occasionally found it difficult to hear /*h*/, and sometimes have been influenced in my decision by the presence of a corresponding but more audible sound (Karore /*r*/ in some words, Kaulong /*k*/ in others) in cognates from other languages.
13. These correspondences with Psohoh may explain a few odd forms in Sengseng Pidgin, such as *morol* for P. *botol* and *mamkin* for P. *pamkin*. They suggest that between Psohoh and Karore, a combination of nasal plus voiced stop lost the stop. Other Pidgin terms in Sengseng are pronounced with the initial nasal (e.g. *a-mbin bean*).
14. In general, **ŋ* is reflected in Sengseng as /*ŋ*/, but there are a few exceptions in which it appears as /*n*/ (see 4.1). In a number of Kaulong words /*n*/ appears for /*ŋ*/ in the languages east of Sengseng (e.g. Psohoh *e-giŋ bird*) and the /*n*/ may reflect borrowing from Kaulong. The final consonant remains a worry, however.
15. This word is *risik* in Karore. Here, as on the north coast of New Britain there seems to have existed a variant form of POC **tasik* with prenasalisation. See Johnston 1980b:113; Goodenough 1961b. Normally **t* is reflected as an alveolar stop or trill in Sengseng.
16. Blust (personal communication) suggests that, instead, the Sengseng form derived from PMP **ŋipen*. While this source cannot be ruled out, it would represent a deviation from the normal pattern of Sengseng reduction of proto-forms, which rarely results in Sengseng words ending in vowels unless these represent proto-final vowels. I would have expected S. *ŋip* or possibly *ŋipo* (if the final consonant was either dropped in POC or reanalysed in S. as the third person singular possessive suffix).
17. The hyphens separate the article from the noun except in *sa-nuhum*, where *sa* = *tree*.
18. Johnston translates the latter as *later today* (1980a:59). In this particular case, the fact that Lakalai consistently indicates future time in adverbs by using a prefix *ga-* may point to it (or a related language) as the source.

19. There is, however, some evidence that in relatively recent times the Tolai traded down to the Kandrian area in order to obtain shells for their type of money (see Chowning 1978a:200), so that the possibility of a little linguistic influence cannot be ruled out.
20. For my suggestions, based on several considerations, about the order in which the ancestors of various AN languages reached New Britain, see Chowning 1976b:379-380.
21. In the Massim, famous for tabus on the names of dead kin on the father's side (but in Woodlark, surprisingly, reported by Lithgow to be those of one's own clan - 1973:106), the effect varies from nil in Molima, where personal names form a special category, to noticeably influential in Me'udana on Normanby Island (Schlesier 1973:53), and in Wagawaga (Seligman 1910:629). See also Simons 1982:201-203.
22. It was reported in 1981 that the practice of giving penis names has dropped out in the longer-contacted villages.
23. Marriage in Sengseng and Kaulong lasts for eternity, a reason given for traditionally killing a widow when her husband died so that she could not remarry (see Chowning 1980:15-16).
24. Unless differences are specifically mentioned, it should be understood that everything I say about Sengseng custom applies to Kaulong as well.
25. The gulf between the single and the married is very marked in these societies (see Goodale 1980:136-137). It may be that some of the use of the 'married' forms is undertaken just to emphasise the gulf.
26. For this reason, I have not counted married forms on my lists, such as e-mon *bird*, though some of them would have raised the apparent AN content of the basic vocabulary.
27. Unless special payments are made, both sexes have to avoid the graves of certain affines, so men also need to know where the wife's parents and others are buried. A man who marries a kinswoman already has most of this knowledge. The tabus still hold for marriages between kin.
28. Adoption was common because of the large number of orphans resulting from the killing of widows; they might be adopted by the mother's kin after spending early life in the father's village.
29. This objection is particularly pertinent because of the peculiarities of Blust's PSM test list. First, it is culture-bound, as 'basic vocabulary' by definition should not be, because it includes referents to practices which are not universal in the region under consideration; see no.93, *to pound, beat - rice, prepared food*; no.68, *to sew (clothing)*; and no.69, *needle*; no.70, *to shoot (an arrow)*. (It also includes no.126, *lake*, though lakes do not occur in all environments.) The list also discriminates against many OC languages in which a single term is used for related concepts (e.g. *hit - kill, long - far, good - correct*) which are said to have been lexically distinguished "in an ancestral stage of their development" (Blust, personal communication). If the single form is not AN, it is scored as two minuses.
30. Although, as regards subgroupings of specific languages, we agree only on Lamogai, we are still both discussing the same languages.

31. I had been doubtful about attributing this form to direct inheritance in S., because the POC form I had seen was *siŋa(R), probably reflected by the first two syllables of S. siŋanan *day* beside siŋaŋ *sun*.
32. Surprisingly, Johnston has overlooked a derivative of this root in the language he knows best: Lakalai reki *waters tabu to women*.
33. When I wrote that, I was mistaken in not attributing the vowel shifts in Arawe to umlauting, being unaware that the shapes of words for body parts that I collected were affected by the vowel in the suffixed first person singular possessive.
34. Johnston attributes a number of lexical items found in SWNB to the joining of AN and NAN roots.

BIBLIOGRAPHY

BLUST, R.A.

- 1977 A rediscovered Austronesian comparative paradigm. *Oceanic Linguistics* 16/1:1-51.
- 1981a Variation on retention rate among Austronesian languages. Paper read at the Third International Conference on Austronesian Linguistics, Bali. Mimeo.
- 1981b Personal communication.

CAPELL, A.

- 1971 The Austronesian languages of Australian New Guinea. In Sebeok, ed. 1971:240-340.

CHOWNING, A.

- 1966 The languages of southwest New Britain. Paper read at the 11th Pacific Science Congress, Tokyo.
- 1969 The Austronesian languages of New Britain. *PL*, A-21:17-45.
- 1973 Milke's "New Guinea cluster": the evidence from northwest New Britain. *Oceanic Linguistics* 12:189-243.
- 1976b History of research in Austronesian languages: New Britain. In Wurm, ed. 1976:179-195.
- 1976b Austronesian languages: New Britain. In Wurm, ed. 1976:365-386.
- 1977 *An introduction to the peoples and cultures of Melanesia*. Second edition. California: Cummings.
- 1978a Changes in West New Britain trading systems in the twentieth century. In J. Specht and J.P. White, eds *Trade and exchange in Oceania and Australia*. University of Sydney Press.
- 1978b Comparative grammars of five New Britain languages. In Wurm and Carrington, eds 1978:1129-1157.

CHOWNING, A.

- 1980 Culture and biology among the Sengseng. *Journal of the Polynesian Society* 89/1:7-31.
- 1983 Interaction between Pidgin and three West New Britain languages. In *Papers in pidgin and creole linguistics No.3*, 191-206. PL, A-65.

CLARK, R.

- 1979 Language. In J.D. Jennings, ed. *The prehistory of Polynesia*, 249-270. Canberra: ANU Press.

DYEN, I.

- 1965 *A lexicostatistical classification of the Austronesian languages*. *IJAL* Memoir 19.

FRANKLIN, K.J.

- 1977 The Kewa language in culture and society. In Wurm, ed. 1977:5-18.

FRAZER, J.G.

- 1922 *The Golden Bough*. Abridged edition. New York: Macmillan.

GOODALE, J.C.

- 1980 Gender, sexuality and marriage: a Kaulong model of nature and culture. In C. MacCormack and M. Strathern, eds *Nature, culture and gender*, 119-142. Cambridge University Press.

GOODENOUGH, W.H.

- 1961 The Willaumez languages of New Britain. Paper read at the 10th Pacific Science Congress, Honolulu. Mimeo.

GRACE, G.W.

- 1964 Movement of the Malayo-Polynesians 1500 B.C. to A.D. 500: the linguistic evidence. *Current Anthropology* 5:361-368.
- 1971 Languages of the New Hebrides and Solomon Islands. In Sebeok, ed. 1971:341-358.
- 1981 Indirect inheritance and aberrant Melanesian languages. In Hollyman and Pawley, eds 1981:255-268.

HOLLYMAN, J. and A. PAWLEY, eds

- 1981 *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 255-268. Auckland: Linguistic Society of New Zealand.

JAMESON, R.D.

- 1956 Names. In Leach, ed. 1956:781-782.

JOHNSTON, R.L.

- 1980a *Nakanai of New Britain: the grammar of an Oceanic language*. PL, B-70.
- 1980b The languages and communities of the Kimbe Bay region. In Johnston ed. 1980:107-158.
- 1981 The quest for Proto-Southwest New Britain. Paper delivered at the 1981 annual congress of the Linguistic Society of Papua New Guinea.

JOHNSTON, R.L.

- 1983 The continuing quest for Proto-Southwest New Britain. Paper read at the 15th Pacific Science Congress, Dunedin.

JOHNSTON, R.L., ed.

- 1980 *Language, communication and development in New Britain*. Ukarumpa: Summer Institute of Linguistics.

KEESING, R.M. and J. FIFI'I

- 1969 Kwaio word tabooing in its cultural context. *Journal of the Polynesian Society* 78/2:154-177.

LEACH, M., ed.

- 1956 *Standard dictionary of folklore, mythology and legend*. New York: Funk & Wagnall.

LINCOLN, P.

- 1978 Reef-Santa Cruz as Austronesian. In Wurm and Carrington, eds 1978:929-967.

- 1979 Proto-Oceanic cognate sets: a compilation. Computer printout.

LITHGOW, D.

- 1973 Language change on Woodlark. *Oceania* 44/2:101-108.

LOUKOTKA, C.

- 1957 Classification des langues papoues. *Lingua Posnaniensis* 6:19-83.

LYNCH, J.D.

- 1975 *Lenakel phonology*. *Working Papers in Linguistics, University of Hawaii* 7/1.

- 1981 Melanesian diversity and Polynesian homogeneity: the other side of the coin. *Oceanic Linguistics* 20/2:95-129.

METRAUX, A.

- 1956 Name tabu. In Leach, ed. 1956:782-783.

PAWLEY, A.

- 1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12:103-188.

- 1981 Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In Hollyman and Pawley, eds 1981:269-309.

SCHLESIER, E.

- 1970 *Me'udana (Südost-Neuguinea)*, 1. Die soziale Struktur. Braunschweig: Albert Limbach.

SEBEOK, T.A., ed.

- 1971 *Current trends in linguistics*, vol.8: *Linguistics in Oceania*. The Hague: Mouton.

SELIGMAN, C.G.

- 1970 *The Melanesians of British New Guinea*. Cambridge University Press.

SIMONS, G.F.

- 1982 Word taboo and comparative Austronesian linguistics. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.3: *Accent on variety*, 157-226. PL, C-76.

THROOP, C. and L. THROOP

- 1980 The languages and communities of the Kandrian inland region. In Johnston, ed. 1980, 225-240.

TYLOR, E.

- 1870 *Researches into the early history of mankind*. Second edition. London: John Murray.

WURM, S.A.

- 1975 The East Papuan Phylum in general. In Wurm, ed. 1975:783-804.

WURM, S.A., ed.

- 1975 *New Guinea area languages and language study*, vol.1: *Papuan languages and the New Guinea linguistic scene*. PL, C-38.
- 1976 *New Guinea area languages and language study*, vol.2: *Austronesian languages*. PL, C-39.
- 1977 *New Guinea area languages and language study*, vol.3/1-2: *Language, culture, society, and the modern world*. PL, C-40.

WURM, S.A. and Lois Carrington, eds

- 1978 *Second International Conference on Austronesian Linguistics: proceedings*. PL, C-61.

LANGUAGES OF NORTH AND CENTRAL VANUATU: GROUPS, CHAINS, CLUSTERS AND WAVES

Ross Clark

1. INTRODUCTION

Nowhere in Melanesia are more languages spoken by fewer people than in Vanuatu. Tryon (1976) lists 105 languages among a population of little more than 100,000¹. The absolute degree of diversity is not so high as these figures might suggest, however, since all these languages belong to the Oceanic subgroup of Austronesian, and in fact probably to not more than two first-order subgroups of Oceanic. Three clear divisions have generally been recognised by earlier researchers. The Emae, Fila-Mele and Futuna-Aniwa languages are Polynesian. (See Clark 1978 for their position within the Polynesian family.) The eight languages of the southern islands (Eromanga, Tanna and Aneityum) constitute a Southern Vanuatu Subgroup (Lynch 1978), and are not closely related to any of the others. (But cf. section 9.) This leaves 94 languages spoken from the Banks and Torres islands south to Efate, a region which I will refer to as North and Central Vanuatu (NCV). The unity of these languages as a subgroup, and their internal relations, will be the subject of this paper.

Several linguists have proposed classifications of the NCV languages as a part of larger surveys of Oceanic or Austronesian (Capell 1962, Dyen 1965, Grace 1955). The two most important studies, however, have been by Pawley (1972) and Tryon (1976). Pawley studied a large sample of 'Eastern Oceanic' languages, 15 of which were from NCV — mainly those described by Codrington (1885) and Ray (1926) — and applied classical subgrouping arguments based on innovations in phonology, morphology and lexicon. Tryon collected basic word lists from more than 300 localities in every part of Vanuatu, 179 of which are published in his book. His classification is primarily based on cognate percentages calculated on these lists. In section 3 below I compare Pawley's and Tryon's classifications in detail.

The aim of the present paper is to integrate the results of these two very different studies and see whether any further conclusions can be reached. Pawley's subgrouping arguments need to be checked against the much fuller data now available, and Tryon's vocabularies should yield further results from qualitative as well as quantitative methods. While Pawley and Tryon are my main sources, I have made use of many other descriptions of particular languages, a list of which will be found in Appendix 1.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 199-236. *Pacific Linguistics*, C-88, 1985.

© Ross Clark

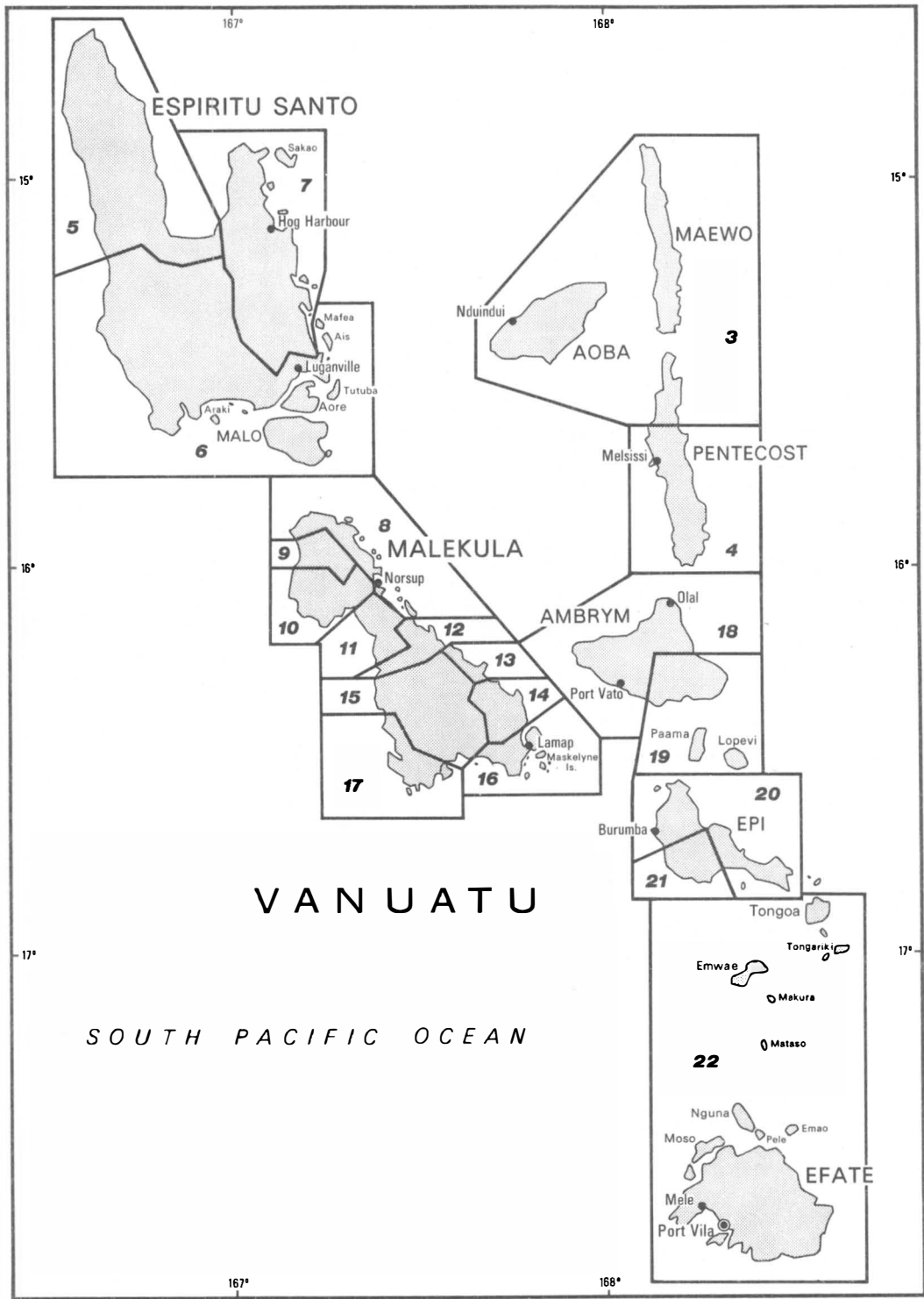


Figure 1 Vanuatu: The 22 local groups

2. THE 22 LOCAL GROUPS

I found it convenient for my own work, and I expect it will also help the reader of this paper, to group 94 languages into a smaller number of working units. Tryon's lexicostatistical figures provided a convenient rough measure of similarity. Tryon considers dialects to belong to the same language if they share more than 80 percent cognates; I therefore chose a figure of 70 percent cognates to define languages which were 'closely related'. A preliminary check showed that several additional connections could be made if this criterial percentage were lowered to 69, so the latter figure is the one used here. Given this criterion, a chain of closely-related languages can be defined as a maximal set of languages, each one of which is connected to every other by a sequence of pairwise closely-related languages (even though pairs of languages at distant points on the chain may not themselves be closely related). This procedure gathered about three quarters of the NCV languages into chains of from two to sixteen languages. The remaining relatively isolated languages (which had no percentage higher than 68 with any other language) were simply grouped with their closest cognate neighbour. The resulting 22 groups are listed in Table 1, and their locations are shown in Fig. 1 (except for groups 1 and 2 which cover, respectively, the Torres and the Banks islands, immediately to the north of the map).²

Although the definition of the groups was made on purely lexicostatistical grounds, all groups are geographically coherent. This is reflected in the names given to them in Table 1. (Some groups with only two members are not named.) The name is followed, in the Table, by a list of those languages, if any, which form a closely-related chain, then by a parenthetical figure which is the minimal connecting percentage within the chain. The relatively isolated languages are then listed (their names being underlined), with their highest cognate percentage. Some groups, such as the Banks (2) thus consist entirely of a chain of closely-related languages; others have in addition an 'outlier' which, though not closely related, has its highest percentage with some language in the chain (e.g. Raga with Baetora in the Aoba-Maewo chain (3)). Still others consist entirely (9-12) or largely (4,7,15) of relatively isolated languages, and might be thought of as 'clusters' rather than chains.

It is interesting to compare these groups with those arrived at by Guy (1982), who applied quite different procedures to the same data. With the following minor exceptions, each one of my local groups corresponds to a node in Guy's dendrogram: (1) Groups 1 and 2 together form a unit in Guy's classification, but the subdivision is different; similarly for groups 11 and 12. (2) Group 8 is divided into a northern and a southern part (see note 2). (3) Tolomako is placed in group 6 rather than group 5, and Labo is an isolate within the Malekula group, whereas I assign it to group 17. See Guy's note on p.314 on the tendency of his procedure to distort the position of isolated languages. (4) Guy's sub-classification of group 6 is somewhat different from mine.

Likewise, in both Pawley's and Tryon's subgroupings, with one exception each of my 22 local groups falls entirely within a lowest-order subgroup. The exception is, again, the very isolated Malekula language Labo. I have assigned it to group 17, along with Southwest Bay and Malfaxal, since its highest cognate percentage (54) is with Southwest Bay. Tryon, however, places Labo in the Malekula Interior Group, while Southwest Bay and Malfaxal are in the Malekula Coastal Subgroup — a very high-level separation.

3. PAWLEY'S AND TRYON'S SUBGROUPINGS COMPARED

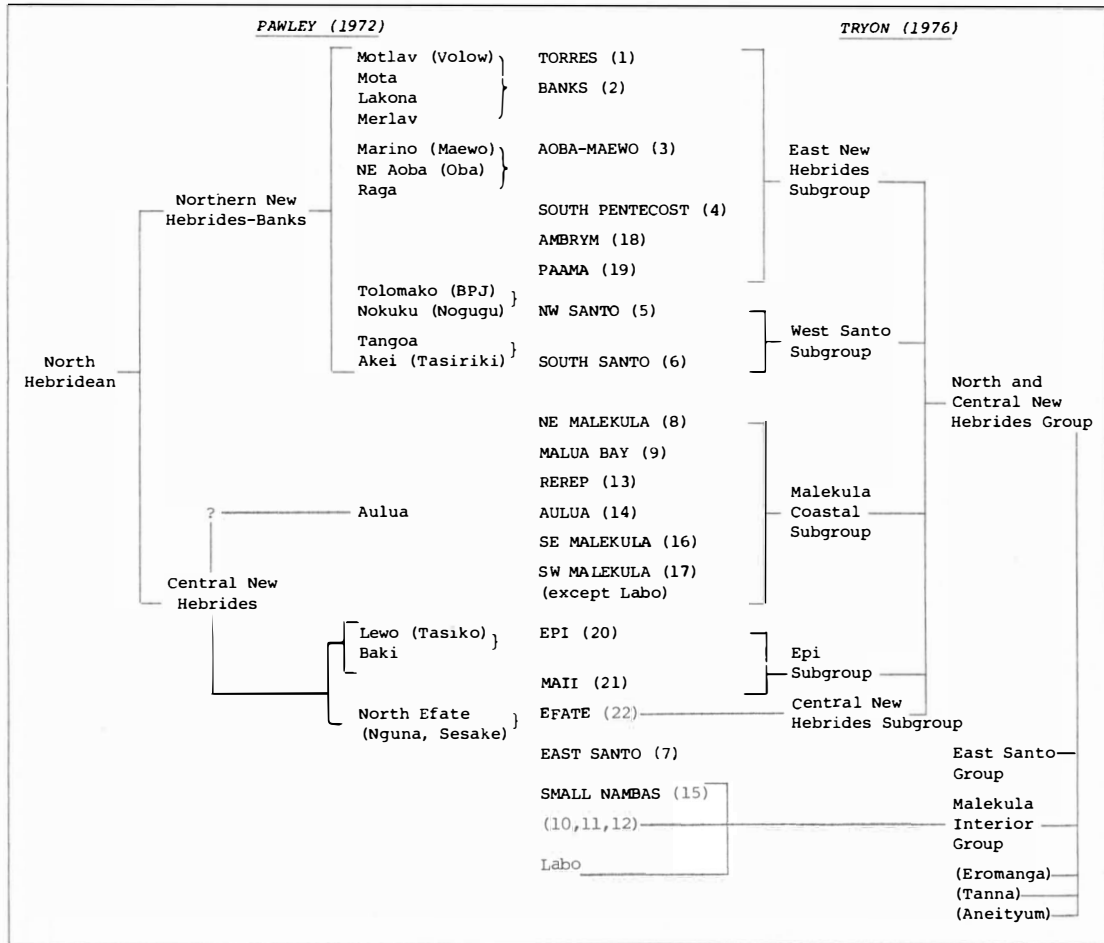


Figure 2. Comparison of Pawley and Tryon classifications

The units just defined make it possible to concisely compare the two earlier classifications of NCV languages. The comparison is shown in Figure 2. Pawley's sample includes languages from only seven of the twenty-two local groups, so that much of what Tryon proposes has no counterpart in Pawley's study. Still, where the two can be compared, they agree more often than not, which is significant when one considers the very different data bases and methods on which their conclusions were based. The following four points are common to Pawley and Tryon:

- There is a single large subgroup (Pawley's North Hebridean, Tryon's North and Central New Hebrides) which includes the great majority of NCV languages and no languages from outside NCV.
- The Torres, Banks and Aoba-Maewo-Raga languages are closely related.
- Most of the languages of Santo are similarly closely related.
- These two groups are separated at a fairly high level from those of Malekula, Epi and Efate.

The only point on which the two subgroupings could actually be said to disagree is that Tryon's NCNH splits into five co-ordinate subgroups, whereas Pawley hypothesises a binary split into Northern New Hebrides-Banks (which is not further divided) and Central New Hebrides, which consists of Epi-Efate and a speculative Malekula group (represented by only one language).

4. NOTES ON THE PRESENT STUDY

My approach in this paper will be similar to Pawley's, in that I look for phonological, lexical and grammatical innovations as evidence of a shared history among languages. I re-examine all of Pawley's subgrouping criteria, confirming some and rejecting or modifying others on the basis of the more abundant data now available, and introduce some new evidence.

The base line for determining what is an 'innovation' at the higher levels, is Proto-Oceanic (PO), for which published sources provide a large number of lexical reconstructions (see Wurm and Wilson 1975) and a fairly clear outline of the grammar (Pawley 1972, 1973). At lower levels, I have used some of my own reconstructions of Proto-North and Central Vanuatu (PNCV). See section 4.2 for more details.

4.1 Diffusion

Tryon claims that in his study "the number of borrowings, by the very nature of the basic wordlists employed, is potentially small and should not have much effect on the subgrouping" (1976:76). I must dissent strongly from this opinion. Although the historical phonology of most NCV languages is not yet well enough known to enable us to detect borrowings, the Polynesian lists show the following (minimum) percentages of words borrowed from their neighbours: Mele-Fila 10%, Emae 7%, Futuna-Aniwa 4-5%. These are basic vocabulary items, so it is not possible to explain them as resulting from cultural or environmental novelties as perceived by Polynesian immigrants. Grammatical influence has also been considerable (Clark 1984). Guy (1982) points out that without the extraneous knowledge of their Polynesian origin, Tryon's data would have led to their classification as a somewhat aberrant group most closely affiliated to the Efate-Shepherds languages (group 22).

Thus lexicostatistical data alone cannot distinguish between a historically intrusive group which has become naturalised through prolonged diffusion (such as the Polynesian outliers) and an autochthonous group which has simply been highly innovative (as is apparently the case with the East Santo cluster). Borrowing of both core vocabulary and grammatical features must be accepted as probably endemic, and capable of distorting the picture of purely genetic affiliations.

4.2 Comparative phonology

In the next section I will present evidence for the existence of a NCV subgroup, and with the reservations explained above I will assume the existence of a corresponding proto-language (PNCV). Forms cited as PNCV are mainly based on my own work (Clark n.d.), and my understanding of the phonological history of these languages is in most cases very sketchy.

The tables given by Tryon (1976:11-50), showing correspondences between PO and his 179 dialects, are useful for an overview, but contain a considerable number of errors in addition to the inevitable uncertainties. A thorough revision would be far beyond the scope of this paper, and probably a waste of time until more and better data become available. However, Table 3 gives what I believe is a more satisfactory list of consonant correspondences between PO, PNCV, and a selected group of languages which will be most frequently referred to below.³

Some general features of the sound correspondences should be noted. PNCV *ʔ is lost in all languages except Namakura (22), hence its presence is often indeterminate if neither a Namakura reflex nor external witnesses are available. I have not indicated this indeterminacy in my reconstructions. The contrast between PNCV *d and *nr is clearly preserved in South Efate (Clark 1985); whether any other languages reflect it is not yet certain. I have used *D for a consonant indeterminate between *d and *nr. Blust (1978) has argued that some Epi languages have distinct reflexes of PO *n and *ñ, but this too will often be indeterminate, and I write *n unless there is positive evidence for *ñ. PNCV *y has few if any overt segmental reflexes, but seems to be reconstructible in a few items on the basis of its effect on surrounding vowels (Clark 1985).

The inclusion of PNCV *R in the table merely points to a complex and still unsolved problem. Pawley (1972:30) noted that PO *R was commonly retained in Mota and other Banks-Torres languages, but lost in NCV languages further south. As *R is also lost in Central Pacific and Nuclear Micronesian languages, he saw this a potential subgrouping diagnostic. However, as Geraghty (1978) has observed, Tryon's data show a much more complex situation in NCV than had previously been apparent. *R appears to be lost in all NCV languages in a final syllable before *a (e.g. PNCV *memea *red* < PO *meRa, *via *taro sp.* < PO *piRa). In other environments, the Banks and Torres languages generally preserve *R; the remaining languages either lose it or preserve it, with no obvious phonological conditioning: compare North Efate nearu *Casuarina* (< PO *yaRu), na-sakau *reef* (< PO *sakaRu).⁴ Although the boundary between zero and non-zero reflexes of *R most typically lies between the Banks-Torres languages and the rest, there are many instances of displacement north or south of this line. For instance, only the Torres group retains *R in *ŋaRi *almond* (Toga ger, but Mota gai), whereas in *paRi *stingray*, non-zero reflexes appear not only throughout Banks-Torres but also in groups 3,4,5 and 18 (Nduindui vare, Seke kofer, Tolomako vari, North Ambrym kenen-ver).

Grace (1976) has observed that some NCV languages show reflexes of PO final consonants, sometimes with a following vowel. Items I have noted with retained final consonants are *uRat *vein* (T31), *kulit *skin* (T39), *qudaŋ *crayfish* (T62), *manuk *bird* (T68), *ñamuk *mosquito* (T69), *kawaR *root* (T105), *pilak *lightning* (T122), *saqat *bad* (T164), *taqun *year* (T209), and *tuqud *stand* (T234). Distribution of retained finals varies greatly from item to item, and does not define a single area, though the concentration is highest in groups 6-10 (Santo and north Malekula). The final in *crayfish*, for example, is retained in only a handful of scattered languages, whereas many reflexes of PNCV *saʔati *bad* occur from the Banks to south Malekula.

Comparative phonology proves to be a disappointment in the search for diagnostic innovations for subgroups of NCV languages. Most of the sound changes apparent on the basis of presently available data are either too commonplace or too restricted in extent to be of interest. For example,

as mentioned above, all NCV languages except Namakura lose PNCV *ʔ (< PO *q), but there is no other reason to place Namakura in a primary subgroup against all the other languages. Glottal stop, as elsewhere, is simply chronically liable to loss. On the other hand, the shifts of PNCV *k > s and *l > c are quite unusual, but are restricted to the Bierebo and Baki languages of Epi (20), and hence of no interest except for the detailed comparative study of the Epi languages. There are a few changes however, that encompass several languages and are phonetically interesting enough to warrant some discussion.

4.2.1 The apico-labial shift

Many languages of Santo and north Malekula show evidence of what I will refer to as the *apico-labial shift* – that is, they reflect the PNCV labial consonants *v, *b and *m either as the corresponding dentals, or as apico-labials (consonants articulated with the tongue between the lips).⁵ The following languages show evidence of this change (those in capitals have the apico-labial articulation):

Group	Languages
5	Tolomako
6 (a)	Roria
6 (b)	TANGOA, ARAKI
6 (d)	MAFEA, AORE, Tambotalo
7	Butmas-Tur, Lore diakarkar/Shark Bay, Sakao
8	VAO, MPOTOVORO, Vovo
9	Mae
10	BIG NAMBAS

This shift is phonetically so unusual, and the languages manifesting it so clearly concentrated in a single area, that we can hardly imagine it not to have had a single origin. Yet it cuts across six different local groups, without including any of them entirely, and thus seems to contradict even our most plausible working assumptions about subgrouping. This is less disturbing, however, if we consider the nature of the change more closely.

The apico-labial shift presumably begins with a shift from labial to apico-labial. This change has two important properties. First, it is reversible (no mergers are involved). Second, it creates a highly-marked type of articulation, very rare in human languages. This means that there will be a high likelihood of subsequent elimination of this series of consonants by further sound change. One possibility is to merge the apico-labials with the other apicals (i.e. dentals). A second is to return them to ordinary labials, thus erasing all evidence of the shift. It is quite possible, then, that all the languages of groups 5-10 originally underwent the apico-labial shift. A majority of languages subsequently reversed the shift and now show no evidence of it (at least from our limited data). The minority listed above either merged the apico-labials with the dentals or preserved the apico-labial articulation.

Does this sound change therefore provide evidence for a large subgroup including all of Santo and northern Malekula? Not necessarily. Camden (1979: 113) makes the interesting observation that in Tangoa apico-labials are a feature of men's speech, whereas women and children use the labials. Apico-labials are a highly salient but very superficial feature of language, and, at least in this case, serve a clear sociolinguistic function. One would expect,

then, that the change could be quite readily borrowed across language boundaries. In the light of this probability, it seems to me that this change, despite its intrinsic interest, is of little value in supporting any subgrouping in this area.

4.2.2 Liquid merger

A distinctive sequence of changes takes place in a considerable number of languages in the central area – south-east Malekula, Ambrym, Paama and Epi. The changes involved are (1) merger of PNCV *r and *l, (2) loss of the resulting liquid in some environments, most commonly before a, and (3) conditional shift of *t to r, which may or may not merge with the liquid. The following chart summarises the participation of various languages in this series of changes:

Group	Language	*r/l merge	liquid lost	*t > r
13	Rerep	+	-	-
	Unua	+	-	-
14	Burmbar	+	-	-
	Aulua	-	-	-
16	Pt Sandwich	+	+	+
	Axamb-1	+	+	+
	Axamb-2	+	()	+
	Maskelynes	+	-	-
18	North Ambrym	-	-	-
	Lonwolwol	-	-	-
	Dakaka	+	+	+
	Port Vato	+	+	+
19	SE Ambrym	+	+	-
	Paama	+	()	-
20	Lewo	+	()	+
	Bierebo	+	()	+
	Baki	+	()	+
21	Maii	+	()	+
	Bieria	+	()	-

Notes: () means that the change takes place in a few items only.

Axamb-1 = Maxbaxo and Avok dialects, Axamb-2 = Axamb dialect. The merged liquid in groups 13 and 14 is r, in groups 18-21 l, and in group 16 r or l according to local phonological conditions.

It will be noted that, like the apico-labial shift, this change cuts through groups of otherwise closely-related languages (groups 14 and 18). The details differ from language to language and even within dialects of the same language, as in the case of Axamb. It does not coincide with any other evidence for a subgroup in the area.

4.2.3 West Santo changes

Two uncommon sound changes are found in a number of West Santo languages of groups 5 and 6. The velar nasal merges with the dental (*g > n), and the prenasalised dental stop *D merges with *q, giving velar or glottal stop reflexes:

Group	Language	*g > n	*d > *q
5	Valpei	-	k
	Nokuku	+	?
	Vunapu	+	?
	Tasmate	+	k
6a	Wusi	±	-
	Malmariv	+	gk
	Navut	+	(g)k
6b	Akei	+	k
	Fortsenal	+	k
	Wailapa	+	-
6c	Morouas	+	-
	Amblong	+	-
	Narango	+	-
6d	Malo	±	-

5. THE NORTH AND CENTRAL VANUATU GROUP

As noted above, Pawley and Tryon agree in assigning the great majority of NCV languages to a single subgroup. In Tryon's study, two small sets of languages — East Santo (ES) and Malekula Interior (MI) — are excluded from the larger group on the basis of low cognate percentages with the other NCV languages. If this classification is interpreted as a family tree, it implies that the remaining languages constituted a subgroup after their separation from East Santo and Malekula Interior. Evidence of such a division would be innovations shared by these majority languages but not by ES or MI languages — to put it in more traditional terms, 'archaic features' in the latter. No such evidence has been found in this study. On the contrary, the ES and MI languages share not only a number of innovations of NCV as a whole, but some restricted to subgroups within NCV, such as Santo or Central Vanuatu. What this suggests is that the deviance of the ES and MI groups results not from an early separation, but from a relatively rapid rate of innovation.

This section, then, presents evidence for a NCV subgroup in the form of innovations putatively shared by all of the 94 languages. Items 1-6 are a review of Pawley's evidence for the North Hebridean subgroup. The remainder of the section presents additional evidence, mainly lexical, from my own work.

5.1 Grammatical evidence

1. PO *-mu, second person singular possessive suffix, becomes PNCV *-mwa.

Mota (2), Lewo (20), North Efate (22) -mwa, Baetora (3) -gwa.

There are two difficulties with this item. First, the two languages of Aoba (3) have -mu, evidently preserving the original form. Second, a majority of the remaining languages have dropped the final vowel of the suffix, leaving simply a nasal consonant. The evidence for the change (or lack of it) is thus reduced to the difference between a labial and a labiovelar nasal, a distinction which has been lost in some languages, and which cannot be reliably extracted from most available descriptions, particularly in final position. Despite these reservations, the wide geographical spread of the clearly

innovative languages listed above suggests that this was indeed a change in progress during the late PNCV period, which failed to establish itself in some of the more conservative dialects.

2. PO *koe, second person singular independent pronoun, becomes PNCV *ni(kq)o

Mota (2) i-niko, Atchin (8) i-nik, North Efate (22) niigo.

Pronouns of this form are by no means universal in NCV, but they are the most widespread type, occurring in all major areas. The innovation consists in the prefixation of *ni- and (possibly) the shift from oral to nasal grade of the stop. Deletion of -e cannot legitimately be included, since *ko forms are as widespread as *koe forms in Oceanic and there is no evidence of a *koe antecedent of the PNCV form.

3. PO *(η)ku, first person singular subject pronoun, is replaced by PNCV *na.

Mota (2), Raga (3), Baki (30) na, Aulua (14) ne.

Pawley notes that *na is evidently a reduced form of the independent pronoun *nau, and that a similar development has taken place in the South-East Solomons. Of the NCV languages for which grammatical data are available, an overwhelming majority have lpsg subject pronouns of the form n(V), na being the most common form. None of the pronoun forms suggest a retention of *(η)ku. This innovation therefore seems quite well supported.

4. PO *i, locative preposition, becomes PNCV *a.

Mota (2), NE Aoba (3), Big Nambas (10) a.

PNCV *a survives as a preposition by itself in only a few languages, such as those mentioned above. In many others, however, it can be found in fossil form, as in the word for *where?* (T203), replacing PO *i-pai: Lo (1) a/ve, Malo (6) a/be, Lingarak (12) a/bi, Port Sandwich (16) a/bi, Baki (20) a/be. A few languages have e or i instead of a, but they are not generally among the more phonologically conservative, so that these forms may well be secondary developments.

5. PO *lalo *inside* becomes PNCV *lolo.

Merlav (2), NE Aoba (3), Tolomako (5), Tangoa (6), Atchin (8), Lonwolwol (18) lolo.

A possible counterexample is Rerep (13) raro-. More importantly, the force of this piece of evidence is weakened by the intrinsic high probability of an assimilation of this sort.

6. PO *i could occur in simple constructions of the form *i NP, with a general locative sense, or in more complex structures involving specific locative nouns such as *lalo *inside*, *papo *under*, etc. (Pawley 1972:33, 37, 43). In PNCV it appears that the originally complex structure *a lolo ... NP with the specific meaning of *inside* may have been tending to be used in a more general sense, with the sequence *a lolo reinterpreted as a simple locative preposition and concomitantly reduced in form to *(a)lo. Thus Mota (2) alo vanua ilone *in that place*, Raga (3) la gatava *at the doorway*, Sakao (7) l-uevyoel *in the men's house*, Big Nambas (10) al pitvet *in the garden*, Lewo (20) lo-yumwa *in the house*. Note however that in some languages *alo coexists with the original *a, and also that a parallel generalisation appears to have taken place in the languages of New Ireland (Malcolm Ross, personal communication).

Pawley mentions two other apparent innovations in NCV: the appearance of a post-verbal particle *qalo/*balo *up*, and the shift of *m to *mw in certain lexical items. Both of these are supported by only fragmentary and inconsistent evidence, so I say no more about them here.

7. Biposed negative constructions (those involving both a prefix or preposed particle and a suffix or postposed particle) are widespread in NCV languages, though rare elsewhere in Oceanic. I find such structures in groups 2, 3, 8, 13, 17, 19, 20, 21 and 22. This is about half of the groups for which grammatical information is available, and covers a wide area of the NCV region. The forms of the negative markers are quite variable, but some recurrent traits suggest possible reconstructions. The pre-verbal marker most often reflects PNCV *(st)a(vb)V: Raga (3) hav, Sa (4) tapo, Nokuku (5) sap, Tangoa (6) sopo, Sakao (7) yav, Port Sandwich (16) sba, South Efate (22) tab. The post-verbal element can probably be reconstructed as PNCV *tea: Motlav (2) te, NE Aoba (3) tea, Atchin (8) te, Paama (19) tei, Lewo (20) re, Bieria (21) se.

We might speculate about the earlier history of this system. PNCV *(st)a(vb)V seems to connect with other Oceanic pre-verbal negative markers of the form *ta-, e.g. PPN *ta'e. *tea, on the other hand, recalls PNCV *tea *one*, and may therefore have been originally an optional emphatic after a negative verb (*not one!*), gradually losing its emphatic force and becoming a routine partner of the original negative marker. Compare the parallel histories of *ne ... pas* in French and *ne ... not* in Middle English.

5.2 Lexical evidence

Apparent lexical innovations of PNCV are listed here in summary form. Items with a T-number in parentheses following the gloss are in Tryon's basic vocabulary, and supporting evidence may be found in Tryon 1976. Data for the other proposed innovations are given in Appendix 2.

1. *NECK* (T14), *THROAT*, *VOICE*: PNCV *Dale'o, from PO *leqo *voice, throat*, with prefix of unknown origin.

2. *RIGHT HAND* (T28): PNCV *matu'a, by metathesis from PO *mataqu.

3. *KNEE* (T33): PNCV *bwau-X replaces PO *tudu. The PNCV form evidently represents a compound meaning *head of leg* – compare PNCV *bwatu *head*, which has irregularly lost its medial consonant in this compound. In some languages it is followed by the word for *leg*, which is here represented by X, since there is some uncertainty about the PNCV form.

4. *FLYING FOX* (T28): PNCV *qarai replaces PO *mpenka. The two languages of group 21, however, appear to retain the PO form.

5. *TURTLE* (T85): PNCV *'avua replaces PO *poñu. North Efate (22) vonu is not a direct retention of the PO form, but a borrowing from Polynesian (Clark 1984).

6. *BREADFRUIT* (T100): PNCV *batavu. PO *kulu is not attested in NCV. However, a number of northern NCV languages reflect a PNCV *baeko, which on the basis of numerous Solomons cognates appears to represent a second PO form for *breadfruit*, *mpaReko. The relation between *mpaReko and *kulu is not clear.

7. *ROOT* (T105): PNCV *kawa(Ri), by metathesis from PO *wakaR.

8. *KNIFE* (T147), *CUT* (T249): PNCV *ziba. cf. PO *pansi *split, cut*, *tampak *cut*.

9. *GIVE* (T222): PNCV *lavi. cf. PO *lapi *take from*, *alap *take, touch*. The PNCV form also occurs widely with the meanings *take, bring, carry*, but the extension to *give* appears to be a unique innovation.

The following innovative forms are widespread enough in NCV that they are unlikely to be assignable to any lower subgroup, but they coexist with reflexes of established PO forms.

10. *COCONUT* (T97, 98): PNCV *matu(ki) alongside *niu (PO *niu).

11. *MOON* (T108): PNCV *kabati(ao) alongside *vula (PO *pulan).

12. *STAR* (T109): PNCV *mwasoyo alongside *vitu?u (PO *pituqu).

A possible earlier sense of this word is suggested by Mota mwasoe *disc, planet, morning or evening star*, and by apparent reflexes of PNCV *mwaso meaning *sun* in various Santo languages.

13. *EARTHQUAKE* (T125): PNCV *muki alongside *ruru. The latter appears to derive from PO *dudu *shake* — cf. also Arosi, Kwaio *nuu*, Yapese *durrug earthquake*.

The next few items are not on Tryon's lists but can be compared with established PO reconstructions.

14. *EEL*: PNCV *maraya replaces PO *ntuna.

15. *NETTLE TREE*: PNCV *qalato, from PO *(sa)latoŋ, with initial syllable unattested elsewhere.

16. *CORDYLINE SP.*: PNCV *(qk)aria replaces PO *siRi.

17. *KAVA*: PNCV *maloku replaces PO *kawa (Pawley 1977).

18. *CHEW, REFUSE OF CHEWING (KAVA, SUGAR CANE, ETC.)*: PNCV *samwa, irregular from PO *samuk. Compare the change *-mu >> *-mwa in the possessive suffix (section 5.1).

The final group of PNCV lexical items cannot be compared with established PO reconstructions, but are included because they are well attested in NCV languages and have no known cognates elsewhere.

19. *TOMORROW* (T207): PNCV *marani. cf. PO *dani, PNCV *rani *day*.

20. *AGAINST*: PNCV *(qk)oro. This functions as the second element in compound verbs indicating action which obstruct covers, surrounds, prevents, etc. Pawley (1972, 1977) connects this with other Oceanic forms meaning *fence, enclosure, cut around* etc., but the particular use here appears to be restricted to NCV.

21. *LAPLAP (PUDDING)*: PNCV *loqo. The national food of Vanuatu, a baked pudding made of grated starch (yam, manioc, banana, etc.) with coconut cream and other ingredients. cf. PO *loŋku *bend, fold, roll up*.

22. *CYCAS PALM*: PNCV *mwele.

23. *GHOST*: PNCV *(a)tamate. cf. *?ata *person*, *mate *dead*.

24. *PEACE*: PNCV *tamwat(ae).

25. *CHIEF, BIG MAN, GRADED SOCIETY*: PNCV *subwe. Capell has suggested a connection with PAN *sembaq *worship, honour*.

6. THE NORTHERN AREA

6.1 Evidence for a North Vanuatu group

Pawley's proposed "Northern New Hebrides-Banks" subgroup covers all of the Torres and Banks groups, Aoba, Maewo, north Pentecost and Santo, i.e. local groups 1-3 and 5-7 in the present study — though his sample included only eleven languages, representing groups 2, 3, 5 and 6. Pawley's arguments for this subgroup, on re-examination, are promising but not convincing. This area includes some of the most conservative languages in NCV, and much of their similarity appears to be due to common retention rather than innovation.

6.1.1 Grammatical evidence

1. Languages of this area have a possessive classifier of the form *bula-, glossed by Pawley as *animal property* or *household property*, elsewhere as *prized possession*, the prototypical example being a pig. There is a possibly cognate form in Vao (8) tala- "used for general objects ... but less frequently and in a more personal sense" (Layard 1942:760). (See section 4.2 on apico-labial shift.) As will appear below, languages of northern Malekula (groups 8-10) share many features with those of south Santo. Aside from this Vao form, such a classifier is not known elsewhere.

2. *tamwa *how?* is reflected in Mota (2), Nokuku (5), Tangoa and Akei (6).

3. *tari *many, large number*. Pawley gives reflexes from Mota, Merlav, Lakona and Motlav (2), Marino (3), Nokuku and Tolomako (5), Malo and Tangoa (6). There are additional cognates in Sakao (7) ter, and in neighbouring groups outside NV: Sa (4) and Uripiv (8) tar.

4. Words for *today* are etymologically very diverse in NCV languages, but forms reflecting *bwariki occur throughout groups 2 and 3, and in Vao (8) barigh. There are three possible but problematic cognates from group 6: Malo baridi, Tangoa na-kerkerighi, and Akei ereRi?i (*now*).

The remaining grammatical features discussed by Pawley are in fact restricted to the north-eastern area of NV. I return to them in section 6.2.

6.1.2 Lexical evidence

Pawley lists a dozen candidates for NV lexical innovations. His data, however, are drawn almost entirely from the comparative wordlists in Codrington (1885:36-100), which are strongly concentrated in areas where Codrington worked. More than half of the Vanuatu lists are from the Banks and Torres islands (groups 1 and 2); only one Santo language is included, and none from Malekula. With the much more complete data now available, it appears that these innovations by no means define a coherent NV area. In fact hardly any two of them agree in their extent. They range from *lama *sea*, apparently restricted to the Banks, Torres and north Maewo, to *matu(ki) *coconut*, which occurs over most of the NCV area (see section 5.2.) Only *vatali *banana* comes close to exactly covering the proposed NV area. See Appendix 3 for representative data.

6.2 The north-east

Two grammatical innovations may be restricted to the north-eastern area, groups 1-3 (Torres, Banks, Aoba-Maewo-Raga).

1. The 'independent noun' suffix **(k)i* appears on nouns which normally take a possessive suffix, when no possessor is specified. It is well attested in groups 1-3, but, as Pawley notes, no such suffix is mentioned by Ray (1926) for any of the four Santo languages he describes. Pawley cites some possible examples from Tolomako (5): *matu/i coconut*, *teta/i father* and *namu/gi mosquito*. The last is not likely to be a suffixed noun, and *-gi* here clearly represents a retained final consonant with supporting vowel (see section 4.2). Although *tetai* is semantically in the right area, it appears to be a vocative form which coexists with suffixed *tama-* in referential use. Apparent cognates occur outside the NV area, e.g. Labo (17) *tatai*. Finally, *matu-* may well be a suffixed noun, but if so the *-i* could as well be a relic of the possessive connective PO **qi* (Pawley 1972:35).

2. The conjunction *si or* is found in almost all languages of groups 1-3, and possibly in 4 (Apsa sige), but not in Santo. Pawley notes a possible cognate *se* in Fijian.

One further grammatical innovation mentioned by Pawley, the feminine personal article *-ro-*, is restricted to the Banks and Torres languages (groups 1 and 2) and Marino of north Maewo (3).

The only possible lexical innovation I have found in this area is **bwaratu flying fox*, replacing PNCV **qarai*. However, probable cognates occur not only in adjoining group 4, but also in groups 12 and 15 of central Malekula. (Compare the set of innovative faunal terms discussed in section 7.4.)

Evidence of shared innovations in the north-eastern area is thus rather weak. In view of the stronger case for a Santo subgroup, to be discussed in the next section, it may be possible to explain these few features as NV innovations which have subsequently been lost in Santo. The main basis for the perceived close relation of groups 1-3 is clearly a shared conservatism.

6.3 Santo

Several lexical innovations suggest that all the languages of Espiritu Santo (groups 5, 6 and 7) constitute a single subgroup.

1. *TONGUE* (T5): **meme* replaces PNCV **mea*. Also found in several languages of groups 8-11, and surprisingly in Port Sandwich (16).

2. *TOOTH* (T6): PNCV **livo* is replaced in all Santo languages. The most likely Proto-Santo form is **kuDu*, which occurs in almost all of groups 6 and 7 as well as Tolomako (5) and Vao (8). The rest of group 5 reflects **bati*, apparently a PO form for a specific type of tooth, which is also generalised to *tooth* in Nduindui (3) and in the Efate-Shepherds languages (22). cf. Mota *patiu tusk*, *eye-tooth*.

3. *LEFT HAND* (T29): **marau* replaces PNCV **mawiri*.

4. *SPIDER WEB* (T86): **bwara* replaces PNCV **(tk)alawa*. Also in Atchin (8) and NE Aoba (3).

5. *ROPE* (T144): **asi* replaces PNCV **tali* or **ka(r)o*.

6. *RED* (T156): *(kq)ara replaces PNCV *memea. Also in Nduindui (3). Tutuba (6d) memea appears to be a relic, but could be re-borrowed from Aoba.

Here and elsewhere, languages of north Malekula share various features with those of Santo. Although it cannot be demonstrated as yet by means of phonological history, as would be ideal, the sporadic distribution of these agreements suggests diffusion rather than common ancestry.

The apico-labial sound shift, discussed in section 4.2 above, is another innovation common to a large area of Santo and northern Malekula. For the reasons explained there, however, it is at best weak evidence for any proposed subgrouping.

6.4 The position of the East Santo cluster

As already noted, there seems to be no evidence which would place the East Santo languages (group 7) outside the NCV group. They positively reflect not only a number of PNCV innovations, but also several assignable to later stages or subgroups within NCV (see for example *EAR*, *TOOTH*, *LEFT HAND* above). The apparently aberrant position of these languages thus appears to result from a high rate of lexical innovation, parallel to the high phonological innovativeness of these languages, especially Sakao (Guy 1978, 1982).

Do the languages of group 7 constitute a subgroup? Polonambauk and Butmas-Tur share about 70% cognates, while Lorediakarkar and Shark Bay (LSB) are, as already noted, essentially the same language. PBT and LSB are connected by percentages from 50 to 65. Sakao is related to both these groups at the 40-50% level; this is high for Sakao, whose percentages with even neighbouring languages are rarely higher than the 30s.

No grammatical information is available for languages of this group other than Sakao, so I can cite only a few possible shared lexical innovations. Group 7 appears to uniquely share the forms *bili *wing* (PNCV *kaba), *vuriti *bite* (PNCV *kati) and *baqari *liver* (PNCV *ʔate). Sakao and LSB also agree in showing reflexes of *bwoe-mate (*dead pig*) for *meat* (PNCV *visiko), where no PBT forms are given by Tryon. There are also some possible innovations shared by PBT and LSB apart from Sakao (cf. Tryon's division into Sakao and South-East Santo, 1976:87). The following seem to be unique to PBT-LSB: *lisu *nose*, *vili *penis*, *voDo *vulva*, *sok *fish/bird*, *makarati *thunder, lightning*, *maDavek *heavy*, *vok *vomit*. However only in the last case (Sakao lu *vomit*) is Sakao clearly conservative.

7. THE CENTRAL AREA

The islands from Malekula and Pentecost south to Efate will be referred to here as Central Vanuatu (CV). It will be seen from Fig.1 that this is an area of greater diversity, lexically at least, than the northern region. 15 of the 22 local groups are in the central area, 10 of them in Malekula alone.

7.1 Evidence for a Central Vanuatu subgroup

Pawley's sample of languages from the CV area was extremely defective, consisting of a single language from Malekula (Aulua of group 14), two from Epi (Lewo and Baki, group 20) and two dialects of North Efate (Nguna and Sesake, group 22). Although much of his argument needs to be modified or discarded in the light of further data, there remains a certain amount of evidence to support the inclusion of at least a large majority of the CV languages in a single subgroup.

7.1.1 Grammatical evidence

1. Many CV languages show a distinctive alternation in the initial consonants of verbs, whereby reflexes of PNCV nasal-grade consonants (*b, *d, *g, *nr) occur in certain syntactic categories and the corresponding oral-grade consonants (*v, *t, *k, *r) in others, e.g. North Efate *e pano he goes, he went, pwa vano go!* Broadly speaking, the nasal grade is associated with 'realis' categories, and the oral with 'irrealis' (conditional, future, imperative, etc.) (See Lynch 1975, Walsh 1981 for fuller discussion.) Systems of this kind are present in all languages of groups 4, 19, 20, 21 and 22 for which grammatical data is available, as well as in Raga (group 3).

These languages form a band along the eastern side of the CV area, with the apparent exception of group 18 (Ambrym). No alternation pattern of this kind has been described for any Malekula language, but there are scattered forms which could be relics of such an alternation. Pawley notes the Aulua numerals:

e-nrua	<i>two</i>	roku-rua	<i>seven</i>
e-ntil	<i>three</i>	rok-til	<i>eight</i>
e-mbis	<i>four</i>	rok-bis	<i>nine</i>

The following forms from Tryon's lists are also suggestive:

	<i>heavy</i> (T170)	<i>light</i> (= not heavy) (T171)
Mae (9)	i-ndrov	i-a-lov
Maragus (10)	i-ndiv	i-ti-tev
Unmet (10)	i-ndəv	i-te-tev

Rerep (13) also shows the following unexplained alternations (Morton 1891):

me buretin	<i>I speak the truth</i>	furetin	<i>he speaks truth</i>
i-borai	<i>he says</i>	hini forai	<i>he says</i>
me buri ju	<i>it is paid for</i>	ma se fuiri rumb	<i>it is not yet paid</i>

An optimistic reading of this fragmentary evidence could lead to the conclusion that a North Efate type consonant alternation was an innovation of Proto-Central Vanuatu which had ceased to be productive in a number of local groups. Indeed one may still hope that such a system will appear alive and well in one of the many undescribed languages of Malekula. On the other hand, as Lynch (1975) has shown, the alternation arises from the fusion of a 'realis' particle of the form *mV with the initial consonant of a following verb. Since such particles are common throughout NCV, it would not be surprising to find alternations of this type arising independently more than once.

2. Pawley observes that the lpsg independent pronoun (PO *inau) has *k-initial forms in many CV languages, e.g. Vovo (8) ghina, Big Nambas (10) kana, Litzlitz (11) xine, Unua (13) xina, Nasarian (15) koenoe, Maskelynes (16) kinau, Lewo (20) kinu, South Efate (22) kineu, all suggesting PCV *(kq)inau.

Nearly as widespread in the CV area are velar-initial second independent pronouns, with forms suggesting PCV *(kq)aiqo, e.g. Apma (4) kik, Unua (13) xai, Maskelynes (16) kaiugku, Malfaxal (17) ghayuqu, Paama (19) keiko, Maii (21) kaikə, South Efate (Eton dialect) (22) kag. (See Tryon's lists 211 and 212 for further examples.) Both of these innovative forms, however, are interdigitated with forms reflecting PNCV *inau and *niqo; compare Burmbar (14) lpsg inau, 2psg xaiugk; North Efate (22) lpsg kinau, 2psg niqo.

Pawley notes the irregular reflection of the postverbal completive particle *tua(i) as sua in Aulua and North Efate. I have not found any other CV languages that share this innovation. Since Paama (19) tuai and Baki (20) rue both reflect the original *t-, it would appear most likely that the Aulua and North Efate forms result from independent developments.

Pawley's other proposed grammatical innovations of CV are actually restricted to the Epi and Efate languages, and will be discussed in section 7.2. Here I add two further innovations which are reflected by most CV languages for which grammatical data are available.

3. There is a copula verb of the form *vei. It is reduced to *ve or *vi in many languages, and *b- and *v-initial forms alternate in languages which have the consonant alternation described above, e.g. Sa (4) e/be, Atchin (8) we, Big Nambas (10) v'i, Rerep (13) fe, Labo (17) vi, Paama (19) hi/vi, Baki (20) ve/mbe, North Efate (22) vej/pei. This verb takes NP complements, and sometimes also possessives and adjectives. The fact that it appears to be in complementary distribution with reflexes of PO *pai *make, do* may suggest its origin. An example cited by Ray (1926:414) shows a use of Tolomako (5) vei in a context which is suggestive of the transition involved: *movei tahonai he becomes well (is made good).*

4. Plurality in nouns is marked by postposing the third plural independent pronoun, or a form which can plausibly be derived from such a pronoun, e.g., Sa (4) atuntun-er *the men*, Southwest Bay (17) nimorot ar *the men*, Baki (20) veru nalo *stones*. This construction does not seem to occur in groups 10, 21 or 22, though North Efate (22) has a postposed plural marker maaga, apparently of non-pronominal origin. Sakao (7) is the only NV language which has a similar structure, though here the plural marker occurs only as a suffix to certain postposed determiners, e.g. ara mam+tr *these pigs*.

7.1.2 Lexical evidence

See Appendix 4 for supporting data.

1. *PIG* (T59): *b(ou)kasi replaces PNCV *boyo (PO *mpoRo).
2. *GRASS* (T104): *mwana(iu) replaces PNCV *valisi (PO *palisi)
3. *RED* (T156): *miala replaces PNCV *mea (PO *meRa).
4. *DRINK* (T251): *minu, from PNCV *inu (PO *inum), perhaps by accretion of a verbal particle. Also in Raga and some southern Maewo dialects.

5. *SPONDIAS DULCIS*: *mali replaces PNCV *ʔusi (PO *quRi).
6. *PIGEON*: *kuiba replaces PNCV *bune (PO *mpune).

7.2 Epi and Efate

The epi and Efate languages (groups 20-22) share a few innovations:

1. Pawley noted similarities among the third person independent pronouns of his four CV languages which suggested common developments. With the much more complete sample provided by Tryon's lists, we can now see a great diversity of forms. Within this profusion, the Epi and Efate pronouns are similar enough to each other and distinct enough from the rest that they probably have a common origin:

Group	3SG	3PL
20 Lewo	naga	nagala
Bierebo	naga(na)	nal, lala
Baki	nai	nalo
21 Maii	ganə	galə
Bieria	gana	niga
22 Namakura	-nini	-niar
North Efate	naae	naara
South Efate	nega	neger

These forms could be plausibly derived from hypothetical Proto-Epi-Efate forms *nagaya *third person singular* and *nagara *third person plural*.

The other grammatical innovations suggested by Pawley are of less value:

2. Both Epi and Efate languages have lost dual and trial independent pronouns. But this seems to have happened also in various other NCV languages, e.g. Marino (3), Sowa (4), Wusi and Malo (6), Big Nambas (10). In fact, as Grace (1976:111-112) implies, repeated abolition and reconstruction of dual and trial pronouns may be endemic in Oceanic languages.

3. The second person plural subject pronoun (PO *(ka)muyu) is reduced to the form *kV. But there is a tendency everywhere to reduce such particles to CV shape. Again, several other NCV languages have followed a parallel path, e.g. Raga (3) ghi, Apma (4) ka, Tangoa (6) ka, Sakao (7) ghi, Atchin (8) ka.

4. The reciprocal prefix, PO *paRi-, reduces to *bi-/vi-. But loss of *R is expected here (see section 4.2), and the further reduction of CVV to CV is hardly unusual; in fact the same reduction has taken place in Wayan and Gilbertese as shown in Pawley's data.

Although none of these innovations in itself is particularly good evidence the conjunction of the three does tend to add some support to the hypothesis of an Epi-Efate subgroup.

Pawley mentions three lexical items which in Codrington's lists seem to be uniquely shared by the Epi and Efate languages: rarua *canoe*, goroi *woman*, and tamoli *man*. These now appear to have been a philological will-o'-the-wisp. They are North Efate words, but do not appear in any of Tryon's 15 Epi dialects. Probably they are contaminations in Codrington's sources, originating from the now extinct Livara dialect of North Efate which was spoken at the south-east end of Epi (Ray 1926:198).

There are, however, at least two clear lexical innovations of Epi-Efate:

1. *TONGUE* (T5): PNCV **mea* changes irregularly to **mena*, perhaps by accretion of the third person singular possessive suffix.
2. *TEN* (T196): PNCV **sagavulu* is replaced by **rua-lima* (*two-five*). This innovation is also found in neighbouring Paama (19).

7.3 Central Central

The remaining area of Central Vanuatu — Malekula, South Pentecost, Ambrym and Paama (groups 4 and 8-19) has three conspicuous uniquely shared lexical items, though only the first can be shown to replace a known reconstructed form. See Appendix 5 for data.

1. *HAND/ARM* (T12): PNCV **lima* is replaced by **vara* (cf. PO **qapaRa* *shoulder*). But **lima* is retained in group 10.
2. *PUT, PLACE, LEAVE*: **linji*. cf. PO **linji* *pour, spill, shed, exude*.
3. *PLACE* (n): **uta*, probably from PO **quta* *land, bush, interior*.

7.4 Ambrym and Pentecost

The close cultural and linguistic connections between South Pentecost and North Ambrym are reflected in a number of local innovations shared by groups 4 and 18. The largest number of these are animal names: **simo(lr)i* *crayfish*, **taliteli* *snake*, **bwaseli* *bird*, **tabwaqan* *mosquito*, **masalo* *fish*, **riri* *squid*, **tomo* *rat*, **buli* *butterfly*, **marit* *eel*. Note also **kul* *coconut* and **visavine* *woman* with a unique infix. These innovations sometimes fail to encompass all of groups 4 and 19, and often extend to neighbouring languages as well: the Paama group, Raga and NE Aoba, and various languages of the east coast of Malekula. The confused pattern of isoglosses suggests cross-language and cross-group diffusion.

7.5 West Malekulan

As with the East Santo languages (see section 6.4), there is no evidence that the 'Malekula Interior Group' of Tryon's classification should be separated from the rest of NCV. The languages of this group share many of the innovations of Central Vanuatu just discussed. Their unity within themselves is less clear than in the case of the East Santo cluster. However, there is a group of lexical innovations which are restricted to Malekula and which occur most frequently in groups 10, 11, 12, 15 and 17. Among the innovative forms are **bulaqu* *bone*, **bwaka* *turtle*, **livakat* *night*, **labut* *rat*, **libak* *dog* and **nitukas* *mosquito*. In addition this group of languages has a unique base **izau-* from which the numerals six-nine are formed. Both the semantic range of the items and their variable distributions are similar to those of the Ambrym-Pentecost area discussed in the preceding section.

This group, which we might term 'West Malekulan', differs from Tryon's 'Malekula Interior' in that it also includes Southwest Bay and Malfaxal of group 17. More importantly, it is seen not as a high-level separate group, but as a focus of a series of innovations.

8. SOME DISTRIBUTIONAL PROBLEMS

Several lexical innovations in the NCV area have clear distribution patterns which are nevertheless difficult to reconcile with the subgrouping picture developed in the preceding sections.

The first group cover a majority of NCV languages in both northern and central Vanuatu, but leave a relic area in the south :

YAM (T91, cf. also T211 *year*): PO *qupi is retained in groups 20-22 and in Paama (19). Most other languages reflect PNCV *Damu.

TARO (T92): PO *talo is retained only in North and South Efate (22) na-tale (with unexplained final vowel). The NV area, as well as South Pentecost and Ambrym, has *bweta, while the remainder of the CV languages have *buaqa.

BANANA (T99): PO *punti is retained only in Marino and Central Maewo (3) and in North and South Efate (22) (North Efate naadi, cf. naasu *bow* < PO *pusu). Most other CV languages reflect *vizi, while NV has *vatali.

SAIL (T142): PO *layaR is retained in groups 19-22, while most other languages reflect PNCV *kabani.

An opposite geographical configuration appears with the items for *FIRE* (T117) and *FIREWOOD* (T134). Groups 1 and 2 (Banks and Torres) preserve the PO forms *api *fire*, *lito *firewood*. Almost all other NCV languages have a single form with both meanings. This common form is *api in a few NV languages (Maewo and NE Aoba (3), Valpei (5), Wusi (6)), but elsewhere it is replaced by *kabu (cf. PO *kampu *burn*). The Banks and Torres appear as a relic area here as they did in the treatment of *R (section 4.2).

8.1 Numerals

A majority of NCV languages have replaced the Proto-Oceanic forms for the numerals from six to nine with morphologically complex additive structures, consisting of a base followed by the numerals from one to four. e.g., Mota *levea-tea six*, *levea-rua seven*, *levea-tol eight*, *levea-vat nine*. The forms of the base are so similar as to suggest strongly that a single innovation was involved, though it is hard to reconstruct a precise PNCV shape for the base. A majority of both NV and CV languages have been affected, but the following languages preserve the PO numerals:

- Group 3: Nduindui, NE Aoba and Raga
- 5: All languages except Tolomako
- 6: Tutuba, Aore, Malo
- 8: All languages
- 9: All languages

Geographically, there is an isolated relic area in north-west Santo, and a larger one comprising facing areas of Malekula, Santo (offshore islands only), Aoba and Pentecost. Recall that the two Aoba languages were likewise a relic area in retaining the possessive suffix *-mu (section 5.1).

8.1 Pronouns

Walsh (1982) has called attention to a widespread pattern in the first and second person non-singular pronouns (T214-216). All three plural pronouns are reconstructed with PO initial *k (PO *kami *first person plural exclusive*, *kinta *first person plural inclusive*, *kamuyu *second person plural*). In many NCV languages, however the inclusive pronoun has the regular reflex of PO/PNCV *k, while the other two have a different consonant, which in general is the reflex of PO *ŋg/PNCV *q.⁶ This is illustrated by Raga kamai (*exclusive*) gida (*inclusive*), kimiū (*second person plural*).

The Raga pattern (Nasal-Oral-Nasal), is reflected in most languages of groups 1-7 and group 18, as well as in Vao (8) and Paama (19). Elsewhere, almost every other possible combination of oral and nasal grades can be found. Walsh interprets these facts as evidence of an innovation shared by Tryon's East New Hebrides and West Santo groups. It seems to me, however, that it could as well as be a retention from PNCV. Suppose that at a pre-PNCV stage, all three pronouns have their initial consonants prenasalised (PNCV *qami *qida, *qamuyu), after which *qida reverts to oral grade, perhaps by dissimilation of two successive prenasalised stops. This gives the Raga pattern, which is retained by quite a few languages. However, pressure of analogy is always strong within pronominal systems. Some languages generalise the nasal grade (e.g. groups 8-11), others the oral (Lakona (2), Aoba (3), SE Ambrym (19), Maii (21)), others produce yet other inconsistencies.

9. NCV AND SOUTH VANUATU

Lynch (1978) argues that the eight languages of Eromanga, Tanna and Aneityum comprise a South Vanuatu (SV) subgroup, and speculates that it may be a first-order subgroup of Oceanic. The latter possibility, of course, can only be established by failure to find evidence for subgrouping SV with any other languages below the Oceanic level. The following resemblances between innovative lexical forms in NCV and SV are given as material for further research into the relationship between these two groups. SV forms are either Proto-SV reconstructions from Lynch 1978, or forms from Sie (Eromanga), Lenakel and Kwamera (Tanna) and Aneityum.

RIGHT HAND (T28): PNCV *matu?a, PSV *mwatu-.

FLYING FOX (T79): PNCV *qarai, PSV *(g)kidai.

TURTLE (T85): PNCV *?avua. SIE na-vu, LEN ia-u, ANT na-hou. Lynch derives these forms from PO *ponu, but I believe a form cognate with *?avua will do at least as well.

ROOT (T105): PNCV *kawa(Ri). ANT ne-cvan, showing the same metathesis from PO *wakaR.

STAR (T109): *mwasoyo. SIE mosi, LEN mahau, ANT in-mojev.

EARTHQUAKE: PNCV *muki, SIE no-miux, LEN mwig, ANT no-moi.

CYCAS PALM: PNCV *mwele. LEN n+-m+l.

GHOST: PNCV *(?a)tamate. LEN iarm+ts, ANT natmas.

TOMORROW: PNCV *marani. SIE mran, ANT imrany.

Grammatically, the NCV innovations in locative prepositions may be compared with SIE ra, LEN le. Note also the biposed negative in LEN is-V-aan.

The final group of lexical items are reconstructed only for subgroups of NCV, but have resemblances in one or more SV languages:

LIVER (T20): *mwabwe, occurring widely in Santo and Malekula. Sie mou, LEN nakan-mop, ANT in-mopo-k. cf. PO *maŋpe *chestnut*.

PIG (T59): PCV *b(ou)kasi. SIE no-mpxahi, LEN pukas, ANT pikad.

FIRE (T117): PNCV(?) *kabu. KWM n-ap, ANT in-xab.

TEN (T196): Epi-Efate *ruaIima. SIE naruolem. Like most NCV languages, the SV languages have re-formed the numerals six to nine as additives on a base of five, but there is no apparent formal agreement.

DRINK (T251): PCV *minu. LEN amnuumw, ANT amony, amnyii.

TONGUE (T5): *lua-mea, found in Merlav, Maewo and several central Malekula languages. Sie ne-luam-, KMW ne-ram.

FINGER: PNV *bisu. LEN p+sp+s, ANT nu-ps. Lynch reconstructs PSV *pot(ie), but as in the case of *TURTLE*, the NV form might do just as well.

10. CONCLUSIONS

Pawley's hypothesis of a North Hebridean subgroup appears supported by additional evidence now available. This group, now called North and Central Vanuatu, comprises all non-Polynesian languages from the Banks and Torres to Efate including the East Santo and Malekula Interior groups. The latter languages, which appeared highly deviant in Tryon's survey, now seem to be innovative local groups, but not separated at a high level from their neighbours.

A relatively gradual differentiation of NCV into regional dialects is suggested by the existence of sub-NCV innovations (section 8) which leave relic areas in the north (Banks and Torres), the south (Efate-Shepherds-Epi) and the centre (Aoba and neighbouring islands). Compare the comments by Pawley (1981) on the NCV area as a sort of hyper-Fiji, where regional languages have further split into chains of closely-related languages.

The major division within NCV appears to be between a northern and a central part, with the boundary running between Santo and Malekula and between Raga and the remainder of Pentecost. This corresponds approximately with the north being of a matrilineal, 'dual organisation' social structure, in which the graded society is generally referred to as *subwe, while in the patrilineal area to the south (Malekula, Ambrym, Epi) it is called *maqi. Peter Crowe (personal communication) has also drawn my attention to a musicological boundary at approximately the same place, with horizontal slit-gongs in the north, upright gongs in the south, and a transitional zone where gongs are set in the ground at a 45-degree angle.

The position of some languages lying near this boundary is still somewhat doubtful. North Malekula languages, particularly groups 8-10, show some anomalous features for members of the Central group, as do those of South Pentecost (group 4). Even Raga (3) is in some ways not at home in the northern region. All that is clear at this stage is that no theory which assumes diffusion to be of negligible importance (section 4.1) is likely to succeed in clarifying their position.

The languages of Santo probably comprise a genetic unit. The remaining NV languages seem to be unified mainly by retentions rather than innovations, though obviously the Banks-Torres group has innovated a good deal, with some influence beyond its boundaries. (See *pei *water*, Tlll, for a particularly clear example.)

Within Central Vanuatu, the major split appears to be between Epi-Efate and the rest. Group 19 (Paama and South-East Ambrym) appears somewhat uncertain in its affinities between these two. Epi and Efate may constitute a subgroup, but the evidence is not very strong. The remaining CV area (Malekula, Ambrym and South Pentecost) shows a number of areas of clear shared innovation, probably at a relatively late (cross-language or cross-group) stage.

There are significant shared lexical innovations between NCV and South Vanuatu. In fact, it is hard to find a clear instance of SV preserving an original feature lost by NCV. We must seriously consider the possibility of a larger Vanuatu subgroup, which splits into NCV and SV; or perhaps even a more intimate relation, considering that SV seems to uniquely share a few innovations with sub-sets of NCV languages.

Table 1: the 22 local groups

1. Torres: Hiw, Toga (69)
2. Banks: Lehali, Lehalurup, Motlav, Mota, Vatrata, Mosina, Koro, Wetamut, Lakona, Merlav (72)
3. Aoba-Maewo: Marino, Central Maewo, Baetora, Northeast Aoba, Nduindui (69), Raga (62)
4. South Pentecost: Sowa, Seke (77), Apma (65), Sa (61)
5. Northwest Santo: Valpei, Nokuku, Tasmate, Vunapu, Piamatsina (74), Tolomakó (68)
6. South Santo: (a) Western: Wusi, Malmariv, Lametin, Navut (74), Roria (60)
(b) Southwestern: Akei, Fortsenal, Wailapa, Araki, Tangoa (74)
(c) South Central: Morouas, Amblong, Narango (76)
(d) Southeastern: Mafea, Tutuba, Aore, Malo (73), Tambotalo (64)
7. East Santo: Polonambauk, Butmas-Tur (75), Lorediakarkar/Shark Bay (65), Sakao (50)
8. Northeast Malekula: Vovo, Vao, Atchin, Uripiv-Wala-Rano (71), Mpotovoro (65)
9. Malua Bay, Mae (68)
10. Big Nambas, Maragus (58)
11. Larevat, Vinmavis, Litzlitz (57)
12. Lingarak, Katbol (59)
13. Rerep, Unua (76)
14. Aulua, Burmbar (72)
15. Small Nambas: Letemboi, Repanbitip (69), Dixon Reef (64), Nasarian (?)
16. Southeast Malekula: Port Sandwich, Axamb, Maskelynes (69)
17. Southwest Malekula: Southwest Bay, Malfaxal (71), Labo (54)
18. Ambrym: North Ambrym, Lonwolwol, Dakaka, Port Vato (73)
19. Paama: Southeast Ambrym, Paama (71)
20. Epi: Lewo, Bierebo, Baki (74)
21. Maii, Bieria (71)
22. Efate-Shepherds: North Efate, South Efate (70), Namakura (60)

Note: Group 6 is a single chain of 16 closely-related languages with two outliers. For convenience it has been broken into four subsets with still closer relations among themselves.

Table 2: Consonant correspondences for representative NCV languages

For multiple reflexes, the following notations are used: X/Y means the reflexes are phonologically conditioned, X-Y means the reflexes occur in different dialects; X=Y means that both reflexes are given in sources but I suspect they are not phonemically distinct; and X,Y means that the basis for the multiple reflexes is not known.

PO	*p	*mp	*ŋp	*t	*nt	*k	*ŋk	*q
PNCV	*v	*b	*bw	*t	*d	*k	*q	*?
1. Hiw	v/w	p	kw	t	t	gh	k	0
2. Mota	v/w	p	pw	t/s	n	gh	k	0
3. Nduindui	v	b	qw	t	d	k	q	0
4. Sa	0	b/p	bw	t/c	d/t	0,k	g/k	0
5. Nokuku	v/w	p	pw/p	t	?	k	?	0
6. Malo	v	b	bw	t	d	x,gh	k	0
Fortsenal	v	p=b	p=b	t	k	0	k	0
7. Sakao	y/0	v/dh	v	dh	r	0	gh	0
8. Atchin	v=w	b=p	bw/p(w)	t/ts=c	r/ts=c	0	k	0
Vao	v/v"	p/p"	b	t/h	r	gh	k	0
9. Malua Bay	v	b/p	b	t/s	r	gh/x	q/k	0
10. Big Nambas	v/v"	p/p"	p	t,0	d(r)	x/0	k	0
11. Vinmavis	v	b/m	b	t/h/s	nt	?,x	q/g	0
12. Lingarak	v	b/mp	b(w)	t/s	d/ns	gh	gk	0
13. Rerep	v	b/mp	bw	t/c	r	x/gh	g/gk	0
14. Aulua	v	b/mp	b	t/s	d/s	x-gh	q/gk	0
15. Dixon Reef	v/p	b/mp	b	t/s	d	k,0	q/gk	0
16. Pt Sandwich	v	b	b	r/ts-c	dr	gh	g/q	0
17. SW Bay	v	mp,b	b	t/s	d	?	gk,q	0
Labo	v/w/p	b	b	t/s	d	0,?,k	q	0
18. Lonwolwol	v/w	b,p	p	t/r,c	n/d	h	k	0
19. SE Ambrym	h	v,w	v	t	r	0	k	0
20. Lewo	v/w/0	p	pw,p	t,s/r	t,s/r	k,0	k	0
Baki	v	b	bw	t/r	t/r	k,s	k,s	0
21. Bieria	v	b,p	b	t,s	t,s,d	k	q	0
22. North Efate	v/w	p	pw	t	d	k	g	0
Namakura	v/w	b	bw	t	d	k	q/g	?

PO	*s	*ns	*m	*ŋm	*n	*n	*ŋ
PNCV	*s	*z	*m	*mw	*n	*ny	*g
1. Hiw	s, t, 0	s	m	mw	n	n	g
2. Mta	s	s	m	mw	n	n	g
3. Ndd	h	s	m	gw	n	n	g
4. Sa	s	s	m	mw	n	n	g
5. Nok	s	ts	m	m	n	n	n
6. Mlo	s	nc	m	m	n	n	g
Fts	s	ts	m	m	n	n	n
7. Sak	h	h	m/n	m/n	n	n	g
8. Atc	s	c-ts	m	mw	n	n	g
Vao	h	s	m/m''	mw	n	n	g
9. Mlb	s	c-ts	m	mw	n	n	g
10. Bgn	0	s	m/m''	m/m''	n	n	n
11. Vmv	s/h	nts	m	m	n	n	g
12. Lgk	s	(n)s	m	mw	n	n	g
13. Rep	s	c	m	mw	n	n	g
14. Aul	s	s	m	m(w)	n	n	g
15. Dxr	s	s	m	mw	n	n	g
16. Psw	s	c	m	m(w)	n	n	g
17. Swb	h	s	m	m	n	n	g
Lab	s, 0	s, 0	m	m(w)	n	n	g
18. Lww	s	s, 0	m	mw	n	n	g/n
19. Sea	s	h	m	m	n	n	g
20. Lew	0	0	m	mw	n	n	g
21. Bie	h	h	m	mw	n	n(y)	g
22. Nef	s	s	m	mw	n	n	g
Nmk	h, s	h, s	m	mw	n	n	g

PO	*d,*R	*nd	*R	*l	*w,*p	*y
PNCV	*r	*nr	*R	*l	*w	*y
1. Hiw	gh	t	gh	y	w	0
2. Mta	r	n	r	l	w	0
3. Ndd	r	d	0	l	kw	0
4. Sa	r	d	0	l	w	0
5. Nok	r	?	0	l	0/w	0
6. Mlo	r	d	0	l	w	0
Fts	r	k	0	l	0	0
7. Sak	r	r	0	l	w	0
8. Atc	r	r	0	l	w/0	0
Vao	r	r	0	l	w	0
9. Mlb	r	r	0	l	w	0
10. Bgn	r	d	0	l	w	0
11. Vmv	r	nr	0	l	w	0
12. Lgk	r	dr	0	l	w	0
13. Rep	r	r	0	l	w	0
14. Aul	r	dr	0	l	w	0
15. Dxr	r	dr	0	l	w	0
16. Psw	0/r/l	dr	0	0/r/l	w	0
17. Swb	r	d	0	l	w	0
Lab	x	r	0	l,dh	w	0
18. Lww	r	d	0	l	w	0
19. Sea	l	r	0	l,0	0	0
20. Lew	l	nd,s	0	l	w	0
Bak	c/l	t/r	0	c/l	w	0
21. Bie	l	t,s,nd	0	l	w	0
22. Nef	r	d	0	l	(w)	0
Nmk	r	d	0	l	0	0

Appendix 1: Sources of data

The lists in Tryon 1976 cover all NCV languages mentioned in this paper. Other sources of data are listed in this appendix. Pawley 1972 draws his data from published sources listed below, chiefly Codrington 1885 (C) and Ray 1926 (R). Two other works which cover a number of languages are Charpentier 1982, which gives lexical data on Rerep (13) and all languages of groups 14-17, and Gowers 1976, which has tree names from throughout Vanuatu. Languages names in parentheses below are those used by the source which differ from Tryon's. Sources preceded by + are scripture translations; full references for these are not given, but most may be found in O'Reilly 1958.

Group

1	Toga	C (Lo)
2	Lehali	C (Norbarbar)
	Motlav	C (also Volow)
	Mota	C Codrington and Palmer 1896, +Bible 1912.
	Vatrata	C (Leon and Sasar, Pak)
	Mosina	C (also Vuras)
	Nume	C (Gog)
	Lakona	C (Lakon)
	Merlav	C
3	Marino	C (Maewo), Ivens 1940-2b (Lotor)
	Baetora	Peter Crowe (field notes)
	NE Aoba	C (Oba, Walurigi), Ivens 1940-2a (Lobaha), Suas
	Nduindui	+Matthew and Mark 1973
	Raga	C (Arag), Ivens 1937-1939 (Lamalanga), Walsh 1966
4	Apma	+Mark and John 1977
	Sa	Tattevin 1929, Elliot 1976
5	Nokuku	R (Nogugu), +John 1946
	Tolomako	C (Marina), R (Bay of Sts Philip and James)
6	Akei	R (Tasiriki), +John 1909, Genesis and Jonah 1912
	Fortsenal	Thomas Ludvigson (field notes)
	Tangoa	R, Annand in Macdonald 1889-1891, Camden 1979
	Malo	Landels in Macdonald 1889-1891, +Selections 1954
7	Sakao	Guy 1974, +Psalms 1949, NT Selections 1959
8	Vao	Layard 1942
	Atchin	Capell and Layard 1980
9	U-W-R	R (Uripiv), +Mark, Luke and Acts 1893-1905
10	Big Nambas	Fox 1979a, b, Corlette 1947
11	Larevat	Deacon 1924
	Vinmavis	Deacon 1924 (Lambumbu)
	Litzlitz	Deacon 1924 (Lagalaga)
13	Rerep	Morton in Macdonald 1889-1891 (Pangkumu)
14	Aulua	R
16	Pt Sandwich	Charpentier 1979
	Maskelynes	R (Kuliviu), +Mark 1906

Group

- 17 SW Bay R (Sinesip), Deacon 1924 (Seniang)
Labo R (Meaun), Deacon 1924 (Mewun)
- 18 N Ambrym C (Ambrym)
Lonwolwol R (Ambrim), Paton 1971, 1973, +Acts 1949
- 19 SE Ambrym Parker 1968, 1970
Paama R, Crowley 1982, +New Testament 1944
- 20 Lewo R (Tasiko), Early MS.
Baki R, Fraser in Macdonald 1889-1891, +Matthew and Mark 1911
Philippians and Thessalonians 1914, Psalms 1914
- 21 Bieria Fraser in Macdonald 1889-1891, Luke 1914
- 22 Namakura Field notes
North Efate C (Fate), R (Nguna), Schütz 1969a, b (Nguna), +Bible
1972, field notes by A.J. Schütz, Ellen Facey and
myself
South Efate Field notes, +Mark 1866, Genesis 1874

Appendix 2: Lexical innovations of PNCV: supporting evidence

14. *EEL*: PNCV *maraya: (2) MTA marea (3) NDD marai (5) TMK narae (6) FTS marai (8) ATC mara (22) NMK mara NEF marae. cf. also (16) PSW marir (18) LWW maret (19) SEA melit, though these may be from PNCV *marita *rope*.
15. *NETTLE TREE*: PNCV *qalato: (2) MTA kalato (3) NEA galato NDD qelato (5) NOK elat (7) SAK gholadh (8) ATC kalat VAO -kalat (15) DXR -qalate (16) MSK -qalat (17) SWB -qalat (18) LWW gelar, gelat.
16. *CORDYLINAE SP.*: PNCV *(qk)aria: (2) MTA karia (6) FTS karia (8) ATC kari (15) DXR -karie (16) PSW kari (17) SWB -ari (22) NMK kari NEF -karie.
17. *KAVA*: PNCV *maloku: (6) FTS maloo (8) VAO maloghe (10) BGN m'aləx (13) REP merox (16) PSW maix (17) MFX -malu (19) PAA malou (22) NMK malok NEF -maloku.
18. *CHEW, REFUSE OF CHEWING*: PNCV *samwa: (2) MTA samwai (7) SAK sama- (10) (10) BGN sama- (13) REP jama- (17) LAB samwe (22) NMK humwa- NEF -samwa.
20. *AGAINST*: PNCV *(qk)oro: (2) MTA ghorō (3) NDD koro BAE ghorō (4) PSA goro (5) NOK -kor TMK goro (6) AKE ?oro- TNG ghorō- (7) SAK ghor (8) VAO ghorō ATC hore (16) MSK kokol (17) SWB qor LAB qoxo (18) LWW goro (19) SEA xole (22) NEF koro.
21. *LAPLAP (PUDDING)*: PNCV *loqo: (2) MTA loko (3) BAE logko NDD -loqo (8) VAO -lok ATC lok (15) LET -laqa (16) MSK -logk (17) MFX -loq (18) LWW lok (19) SEA e-ok PAA -loko (22) NMK log.
22. *CYCAS PALM*: PNCV *mwele: (2) MTA mwele (3) NDD gwele (6) FTS mele (7) SAK oemaol (8) VAO mel ATC mwel (11) VMV -mule (17) SWB mweil- (22) NMK mwai NEF -mwele.

23. *GHOST*: PNCV *(a)tamate: (2) MTA tamate (3) NDD tamate (4) PSA atmat (5) NOK temat (6) FTS tamate (7) SAK edhenm (8) ATC ta-mats (10) BGN tam"a (11) VMV -temah (16) PSW ramac (17) SWB temes (18) LWW temar (19) SEA temaet (22) NEF -atamate.
24. *PEACE*: PNCV *tamwate: (2) MTA tamwata (3) NDD tagwata (6) AKE tamata (8) ATC tamat (13) REP damat (16) PSW ramar *neutral place* (17) SWB -tamate LAB -tamate (18) LWW tamar (*sleep*) *deeply, soundly* (19) SEA tamat PAA tomato (22) NEF tamwate.
25. *CHIEF, BIG MAN, GRADED SOCIETY*: PNCV *subwe: (3) MTA supwe *the club, society* (3) NDD huqwe (8) ATC sup *old man* VAO -hube *title for old man* (16) PSW -sub *high man* (19) SEA sup *chief* PAA asuvo *chief* (20) LEW supwe *king* (22) NEF supwe *image of ancestor, god*.

Appendix 3: Northern innovations proposed by Pawley

These tables show the distribution, by local groups, of 12 NV lexical innovations suggested by Pawley (1972:116-117). For each group, the innovative form cited is from the representative language listed at the left, unless otherwise specified. X means that a conservative form occurs in the group, while - indicates either a lack of data, a third form (i.e. one which is equivocal as to the innovation), or cases where the original form is unknown (e.g. *finger*).

	*vatali <i>banana</i>	*bisu <i>finger</i>	*bw(eo) ro <i>ear</i>
1. HIW	votoi	pus-	X
2. MTA	vetal	pisi-u	pworo-/X
3. NDD	fatali/x	bihu	qwero-
5. NOK	vetoli	-	X
6. MLO	vetai	FTS pisi-	pwero-/X
7. SAK	idhel	-	oevaor-
4. PSA	-	-	X
8. ATC	-	buesh	pora-
9. MLB	-	mbis	mboro-/X
10. BGN	-	pise-n [?]	X
11. VMV	-	-	X
12. LGK	-	KTB soemboe-	X
13. REP	-	mbusumbs umb	X
14. AUL	-	BBR na-mboesmboe-	-
15. DXR	-	-	X
16. PSW	-	mbus-	X
17. SWB	-	-	X
18. LWW	-	-	X
19. SEA	-	PAA haasua- [?]	X
20. LEW	-	pasu <i>thumb</i>	X
21. BIE	-	-	X
22. NEF	X	-	X

	<i>*vi(nrl)u skin</i>	<i>*mazi fish</i>	<i>*karivi rat</i>
1. HIW	X	X	X
2. MTA	vini-/X	VTR mes/X	X
3. NDD	vinu-	BAE mas/X	karivi
5. NOK	X	VNP matsi/X	keriv
6. MLO	WUS vinu-/X	manci	xarivi
7. SAK	SKB viri-	enes	SKB ive [?]
4. PSA	SOW vinu-	X	-
8. ATC	vuelvuelu-	VAO na-m"as	n-ariv
9. MLB	no-vlo	nə-mats	na-gharip
10. BGN	n-il	MGS nə-mets	-
11. VMV	ni-vini-	LVT nə-ments/X	-
12. LGK	-	X	-
13. REP	viri-	X	X
14. AUL	-	X	X
15. DXR	X	X	X
16. PSW	X	X	X
17. SWB	X	-	X
18. LWW	X	-	-
19. SEA	X	X	X
20. LEW	X	X	X
21. BIE	X	X	X
22. NEF	X	X	X

	<i>*sari spear</i>	<i>*l(oi)(dt)o spit</i>	<i>*turi/ai body</i>
1. HIW	-	-	-
2. MTA	isar	MRL lot/X	turiai
3. NDD	hari	lito	turegi
5. NOK	-	lotou	-
6. MLO	sari	lito/X	-
7. SAK	eher	X	-
4. PSA	-	X	-
8. ATC	ne-sar	lutou	-
9. MLB	n-sar	-	-
10. BGN	MGS sar	-	-
11. VMV	-	-	-
12. LGK	-	-	-
13. REP	UNA ne-ser	rut	-
14. AUL	BBR -ser	-	-
15. DXR	-	-	-
16. PSW	-	-	-
17. SWB	-	-	-
18. LWW	-	X	-
19. SEA	-	X	-
20. LEW	-	-	-
21. BIE	-	-	-
22. NEF	-	-	-

		*matu/gi <i>coconut</i>	*taDun <i>person</i>	*lama <i>sea</i>
1.	HIW	matu/gi	-	yamə
2.	MTA	matigh	tanun	lama
3.	NDD	matui	RAG atatu	MNO lama/X
5.	NOK	VLP matui	-	X
6.	MLO	MAF m"atiu	FTS takun	X
7.	SAK	SKB netsi	-	X
4.	PSA	-	atuntun	X
8.	ATC	-	-	X
9.	MLB	-	-	X
10.	BGN	m"etu	-	X
11.	VMV	-	-	X
12.	LGK	-	-	X [?]
13.	REP	-metmet	-	X
14.	AUL	-	-	X
15.	DXR	-mat	-	X
16.	PSW	-maru	-	X
17.	SWB	-metu	-	X
18.	LWW	-	-	X
19.	SEA	maetu	-	X
20.	LEW	maru	-	X
21.	BIE	metoma [?]	-	X
22.	NEF	mwaritou [?]	-	X

Appendix 4: Lexical evidence for Central Vanuatu

These tables follow the same conventions as in Appendix 3, with the addition that in the items not on Tryon's lists, conservative forms are cited in square brackets rather than represented by X.

		*bukasi <i>pig</i>	*mwana(iu) <i>grass</i>	*miala <i>red</i>
1.	HIW	-	X	X
2.	MTA	X	X	X
3.	NDD	X	-	X
5.	NOK	X	X	-
6.	MLO	X	X	-
7.	SAK	X	-	-
4.	PSA	X	X	X
8.	ATC	pua	-	MPT -nial
9.	MLB	bukəs	-	i-mel
10.	BGN	pua	-	i-m"ial
11.	VMV	nu-buah	ni-mwini	i-miali
12.	LGK	-	TBB a-mwanai	i-mial
13.	REP	bue	-	-
14.	AUL	bue	na-məne	miel
15.	DXR	buas	LTB monai	i-miemial
16.	PSW	buas	-	-
17.	SWB	ni-buwes	ni-mwenei	ti-memal
18.	LWW	-	PVO bor/minye	-
19.	SEA	PAA fuas	hus/mwənai	-
20.	LEW	pui	ma/mwini	-
21.	BIE	bukah	lu/mwona	-
22.	NEF	NMK -mbokah	-mwənau	miala

	* <i>minu drink</i>	* <i>mali Spondias</i>	* <i>kuiba pigeon</i>
1. HIW	X	-	-
2. MTA	X	[us]	[pwona]
3. NDD	X/RAG <i>mwinu</i>	[RAG <i>uhi/gai</i>]	
5. NOK	X	[ousi]	-
6. MLO	X	-	-
7. SAK	-	[noe]	-
4. PSA	- <i>mini</i>	-	up
8. ATC	- <i>mini</i>	[UWR <i>na-us</i>]	-
9. MLB	- <i>min</i>	-	-
10. BGN	- <i>m"ene</i>	-	ghup"
11. VMV	- <i>min</i>	-	-
12. LGK	- <i>minio</i>	-	-
13. REP	- <i>min</i>	-	-
14. AUL	- <i>migna</i>	-	-
15. DXR	- <i>mən</i>	-	-
16. PSW	MSK - <i>mueni</i>	mar/kokoc	na-xumb
17. SWB	- <i>min</i>	MFX <i>na-van/malmal</i>	no-oimb
18. LWW	- <i>minu</i>	mel	um
19. SEA	- <i>muni</i>	mael	uip
20. LEW	- <i>muni</i>	melmel	kupa
21. BIE	- <i>mun</i>	-	-
22. NEF	<i>munugi</i>	na-mali	wiipa, NMK kiim

Appendix 5: Three Central-Central innovations

1. *HAND/ARM*: **vara*: (4) PSA *ra-* (8) ATC *wera-* (9) MLB *-vəra-* (11) VMV *-vera-* (12) LGK *-vra-* (13) REP *veru-* (14) AUL *vari-* (15) DXR *-vari* (16) PSW *vea-* (17) SWB *-vara-* (18) LWW *wera-* (19) SEA *heo-*.
2. *PUT, PLACE, LEAVE*: **ligi*: (4) PSA *ligi place* (8) VAO *ligi lead, conduct, accompany* ATC *ligi conduct, ferry* (14) AUL *ligi allow* (16) MSK PSW *rigi put* (18) LWW *ligi put, place, let go* (19) PAA *ligi leave, put*.
3. *PLACE (N)*: **uta*: (8) ATC *ut place, time* UWR *nutu place* (10) BGN *nut place* (14) AUL *nuta place, country* (16) MSK *naut-place* (17) SWB *ne-wut space period, part* (18) LWW *or place, weather* (19) SEA *ut place, area, land* PAA *out place*.

ACKNOWLEDGEMENTS

I would like to thank Bob Blust, John Lynch, Andrew Pawley, Malcolm Ross, and Darrell Tryon for their helpful comments on the conference version of this paper; and Peter Crowe, Robert Early, Ellen Facey, Tom Ludvigson and Al Schütz for allowing me to use unpublished materials.

NOTES

1. The grouping of lists into languages and the language and place names used by Tryon (1976) will be followed in this paper, except that 'Vanuatu' is used for 'New Hebrides'. Thus some more recent toponymic reforms are not reflected here; e.g. 'Aoba' is now officially known as 'Ambae'.
2. I work with Tryon's figures rounded to the nearest one percent. On this basis, Lorediakarkar and Shark Bay (group 7) are dialects of the same language, having a shared cognate percentage of 80.5, rounded to 81. On Tryon's own criteria, Lamenu ought to be separated from the rest of Lewo (20), with which it shares no more than 78.8%; and Lelepa (22) ought to be a separate language from both North Efate (78.9%) and South Efate (72.0%). Nevertheless I continue to work with the dialects as grouped in Tryon's table (1976:87-93).

For simplicity's sake, I have also not taken into account the fact that many of the cognate percentages are based on fewer than 200 comparisons, and clearly inflated as a result. This appears to be important only in the case of group 8, which forms a chain only by virtue of the percentage Rano-Vao 71.1%. If we eliminate this (all of Rano's percentages being inflated), group 8 falls into a northern part (Vao, Vovo, Mpotovoro) and a southern (Atchin, Uripiv-Wala-Rano).

3. With the exception of Proto-Oceanic forms, for which I use the standard orthography, all other forms cited in this paper are in a consistent broad transcription with a minimum of special phonetic symbols. The occasional resulting ambiguity seems acceptable at this stage and level of investigation. *In addition to their normal phonetic values*, letters are used as follows:

b = [mb]	ae = [æ]
d = [nd]	oe = [ö, œ]
q = [ŋg]	ue = [ü]
c = [č]	ao = [ɔ]
j = [j]	e = [ɛ]
v = [β]	o = [ɔ]
th = [θ]	a = [ʌ]
dh = [ð]	pw, mw etc. are labiovelar consonants
sh = [ʃ]	p", m" etc. are apico-labial consonants
z = [ts]	
gh = [ɣ]	
g = [ŋ]	
ny = [ñ]	

4. The non-zero reflex of *R is the same as that of PO *d in all NCV languages.
5. All the languages in question have shifted *m, but a few have not shifted *b or *v.
6. Raga is unusual in that PNCV *q has split, being reflected as q in most lexical items, but as k in a few grammatical phonemes.

BIBLIOGRAPHY

- BLUST, Robert A.
 1978 *The Proto-Oceanic palatals*. Wellington: The Polynesian Society.
- CAMDEN, William G.
 1979 Parallels in structure and lexicon and syntax between New Hebrides Bislama and the South Santo language spoken at Tangoa. *PL*, A-57:51-117.
- CAPELL, A.
 1962 *A linguistic survey of the south-western Pacific*. South Pacific Commission Technical Paper 136. Noumea: South Pacific Commission.
- CAPELL, A. and J. LAYARD
 1980 *Materials in Atchin, Malekula: grammar, vocabulary and texts*. *PL*, D-20.
- CHARPENTIER, Jean-Michel
 1979 *La langue de Port-Sandwich (Nouvelles-Hébrides)*. Langues et civilisations à tradition orale 34. Paris: SELAF.
 1982 *Atlas linguistique du Sud-Malakula (Vanuatu)*. Langues et cultures du Pacifique 2. Paris: SELAF.
- CLARK, Ross
 1978 The New Hebridean Outliers. *PL*, C-61:911-928.
 1986 Linguistic convergence in central Vanuatu. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*. *PL*, C-94.
 1985 The Efate dialects. *Te Reo* 28 (in press).
 n.d. Some Proto-North and Central Vanuatu reconstructions with selected supporting evidence. MS.
- CODRINGTON, R.H.
 1885 *The Melanesian languages*. Oxford: Clarendon Press.
- CORLETTE, E.A.C.
 1947 Langue des Big Nambas: le Mallicolo. *Journal de la Société des Océanistes* 3:57-91.
- CROWLEY, Terry
 1982 *The Paamese language of Vanuatu*. *PL*, B-87.
- DEACON, A.B.
 1934 *Malekula: a vanishing people in the New Hebrides*. London: Routledge.
- DYEN, Isidore
 1965 *A lexicostatistical classification of the Austronesian languages*. Baltimore: Waverly Press.

EARLY, Robert

n.d. Lewo-English dictionary: work in progress. MS.

ELLIOTT, George Robert

1976 A description of Sa, a New Hebridean language. MA thesis, Macquarie University, Sydney.

FOX, G.J.

1979a *Big Nambas grammar*. PL, B-60.

1979b Big Nambas custom texts. *Journal de la Société des Océanistes*. 65:286-293.

GERAGHTY, Paul

1978 Review of Tryon 1976. *Kivung* 11:75-86.

GOWERS, Sheila

1976 *Some common trees of the New Hebrides and their vernacular names*. Vila: Forestry Section, Department of Agriculture.

GRACE, George W.

1955 Subgrouping of Malayo-Polynesian: a report of tentative findings. *American Anthropologist* 57:337-339.

1976 Review of Pawley 1972. *Journal of the Polynesian Society* 85:103-112.

GUY, Jacques B.M.

1974 *A grammar of the northern dialect of Sakao*. PL, B-33.

1978 Proto-North New Hebridean reconstructions. PL, C-61:781-850.

1982 Bases for new methods in glottochronology. PL, C-74:283-314.

IVENS, W.G.

1937- A grammar of the language of Lamalanga, North Raga, New Hebrides.
1979 BSOAS 9:733-763.

1940- A grammar of the language of Lobaha, Lepers' Island, New Hebrides,
1942a Melanesia. BSOAS 10:345-363.

1940 A grammar of the language of Lotora, Maewo, New Hebrides, Melanesia.
1942b BSOAS 10:679-798.

LAYARD, John

1942 *Stone men of Malekula: Vao*. London: Chatto and Windus.

LYNCH, John

1975 Oral/nasal alternation and the realis/irrealis distinction in Oceanic languages. *Oceanic Linguistics* 14:87-99.

1978 Proto-South Hebridean and Proto-Oceanic. PL, C-61:717-779.

MACDONALD, Daniel

1889-91 *South Sea languages*. Melbourne: Public Library and Art Gallery.

MORTON, Alex

1891 Grammar of the language spoken at Pangkumu, Malekula. In Macdonald 1889-91:34-72.

O'REILLY, Patrick

- 1958 *Bibliographie des Nouvelles-Hébrides*. Publications de la Société des Océanistes, No.8. Paris: Musée de l'Homme.

PARKER, Gary J.

- 1968 Southeast Ambrym verb inflection and morphophonemics. *PL*, A-15: 27-40.
 1970 *Southeast Ambrym dictionary*. *PL*, C-17.

PATON, W.F.

- 1971 *Ambrym (Lonwolwol) grammar*. *PL*, B-19.
 1973 *Ambrym (Lonwolwol) dictionary*. *PL*, C-21.

PAWLEY, Andrew

- 1972 On the internal relationships of Eastern Oceanic languages. In R.C. Green and M. Kelly, eds *Studies in Oceanic culture history*, 3:1-141. Pacific Anthropological Records Number 13. Honolulu: Bernice P. Bishop Museum.
 1973 Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12: 103-188.
 1977 Proto-Oceanic reconstructions. MS.
 1981 Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In J. Hollyman and A. Pawley, eds *Studies in Pacific languages & cultures in honour of Bruce Biggs*, 269-309. Auckland: Linguistic Society of New Zealand.

RAY, Sidney Herbert

- 1926 *A comparative study of the Melanesian island languages*. Cambridge University Press.

SCHÜTZ, Albert J.

- 1969a *Nguna texts*. Oceanic Linguistics Special Publication No.4. Honolulu: University of Hawaii Press.
 1969b *Nguna grammar*. Oceanic Linguistics Special Publication No.5. Honolulu: University of Hawaii Press.

SUAS, J.B.

- 1912 Mythes et légendes des indigènes des Nouvelles-Hébrides (Océanie). *Anthropos* 7:33-66.

TATTEVIN, Elie

- 1929- Mythes et légendes du Sud de l'île Pentecôte. *Anthropos* 24:983-1004,
 1931 26:489-512, 863-881.

TRYON, D.T.

- 1976 *New Hebrides languages: an internal classification*. *PL*, C-50.

WALSH, D.S.

- 1966 The phonology and phrase structure of Raxa. MA thesis, University of Auckland.

WALSH, D.S.

1982a The restricted distribution of Raga /k/: starting point for a sub-grouping hypothesis. In R. Carle et al., eds *Gava⁴: Studies in Austronesian languages and cultures dedicated to Hans Kähler*, 237-251. Berlin: Dietrich Reimer.

1982b Variation of verb-initial consonants in some Eastern Oceanic languages. *PL*, C-74:231-242.

WURM, S.A. and B. WILSON

1975 *English finderlist of reconstructions in Austronesian languages. (post-Brandstetter)*. *PL*, C-33.

THE STATUS OF FLORA AND FAUNA GLOSSES THAT HAVE BEEN RECONSTRUCTED FOR PROTO-OCEANIC AND FOR THREE SUB-OCEANIC PROTO-LANGUAGES

D. S. Walsh

0. INTRODUCTION

The data base for this paper consists of flora and fauna glosses, as listed in Wurm and Wilson (1975), that have been reconstructed for Proto-Oceanic (POC), Proto-Eastern Oceanic (PEO), Proto-Malaitan (PML), and Proto-Polynesian (PPN).¹ These essentially etic glosses are considered in terms of what notional rank of category within an essentially etic reconstructed taxonomic hierarchy occurs at what level within the linguistic family tree. Some minimal prerequisites are then suggested for the reconstruction of at least notionally emic glosses which might provide a basis for reconstruction of notionally emic taxonomic hierarchies.

1. THE ETIC STATUS OF EXTANT RECONSTRUCTED GLOSSES

The glosses in the data base² are the result, in the main, of comparison of lexical flora and fauna data that happened³ to be available for various relevant contemporary Oceanic languages. The extent of such data for a given language ranges from the many hundreds of items present in some dictionaries, such as Pukui and Elbert's *Hawaiian dictionary* (1965) and Williams' *A dictionary of the Maori language* (1971), down to the mere handfuls of items present within vocabulary lists of around 200 to 1500 words which are all that is as yet available for most of the Oceanic languages of Melanesia. The quality of the glosses in these data ranges from the botanically or zoologically relatively sophisticated and specific, as is often the case in the above-mentioned Hawaiian and Maori dictionaries, down to the botanically or zoologically naive and relatively non-specific *kind of bird/fish/tree/etc.* type of gloss which predominates in most of the sources.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 237-256. *Pacific Linguistics*, C-88, 1985.

© D. S. Walsh

237

All these glosses, however, have one quality in common — they are essentially *etic*, i.e. they each reflect some classificatory framework, either scientific or lay as the case may be, which has come into being *outside* the primary cultural context of the target language. Consequently, any proto-glosses produced by comparison of these *etic* gloss components of form-plus-meaning resemblances are necessarily also *etic*, as are any hierarchic structurings of such proto-glosses.

2. CORRELATION BETWEEN RECONSTRUCTED TAXONOMIC HIERARCHY AND HISTORICALLY DERIVED HIERARCHY OF LINGUISTIC GROUPINGS

The two logically possible extremes of correlation between a given reconstructed taxonomic hierarchy and a given hierarchic ranking of proto-languages in a subgrouping are: (a) that because more general and inclusive categories and category-labels develop before less general and inclusive ones, the former will tend to be reconstructable for higher-order proto-languages than will the latter; and (b) that because recognition and labelling of more specific categories precedes formulation and labelling of more general categories, the latter will not necessarily be reconstructable for higher-order proto-languages than will the former.

A qualified version of position (a) has been adopted by Brown (1981 and 1982 with specific PPN and sub-PPN reference, and 1984 with wide cross-cultural reference) who proposes (1984:24) an ordered sequence of development of highest-order categories and category-labels for each of the two broad semantic fields under consideration — that of botanical nomenclature and that of zoological nomenclature.

My own expectation on commencing this investigation of the Wurm and Wilson data, an expectation based in part on a reading of Brown 1981 and 1982 and in part on personal experience of PPN and sub-PPN reconstruction (Walsh and Biggs 1966, Biggs, Walsh and Waqa 1970), was that the labelled flora and fauna taxa most likely to be reconstructable for POC would include both higher-order items and culturally salient (in so far as cultural salience can be reputedly inferred for 'proto-cultures') lower-order items, whereas for sub-Oceanic proto-languages the labelled flora and fauna taxa most likely to be reconstructable would include not only what was reconstructable for POC but also lower-order items without marked cultural salience.

3. FLORA AND FAUNA GLOSSES THAT HAVE BEEN RECONSTRUCTED FOR POC, PEO, PML AND PPN

The four proto-languages under consideration have been selected because they are the only ones within the Oceanic group for each of which the data base contains a significant number of reconstructions for both flora and fauna taxa. These four proto-languages are here heuristically regarded as having the following hierarchic standings: PEO is a first-order subordinate of POC; PML is a second-order subordinate of one of the first-order subordinates of PEO; and PPN is a third-order subordinate of the other first-order subordinate of PEO.

The various hierarchically ranked orders of flora and fauna items here referred to (following Berlin, Breedlove and Raven 1974, Hunn 1977 and Brown 1984) are: *unique beginner*, e.g. English *plant, animal; life form*, e.g.

English *tree*, *grass* as *kinds of plant*, and *bird*, *fish* as *kinds of animal*; generic, e.g. English *swallow*, *hawk* as *kinds of bird*; specific, e.g. English *wood swallow*, *Pacific swallow* as *kinds of swallow*, and *sparrow hawk*, *goshawk*, as *kinds of hawk*; and varietal, e.g. English *white-breasted wood swallow*, *black-faced wood swallow* as *kinds of wood swallow*.

The level or levels, labelled 0 to 4 in descending order, at which a given kind of item can occur within the taxonomic hierarchy are as follows (again largely following Berlin, Breedlove and Raven 1974, et al): *Level 0* unique beginner; *Level 1* life form, generic (if not subsumed within an overt or covert⁴ life form); *Level 2* generic, specific (if subsumed within a level 1 generic); *Level 3* specific, varietal (if subsumed within a level 2 specific); and *Level 4* varietal. As the ascription of a given hierarchic rank to a given item has, in this context, to be notional because of the limitations of the date, any item in the listings in sections 3.1 to 3.4 which is followed by (?) may well eventually turn out to be more fittingly located at the level immediately below that at which it has been listed.

Where what is, in effect, a composite or multi-labelled proto-gloss, e.g. *crayfish*, *lobster*, *prawn* or *moss*, *seaweed*, has been associated with a given proto-form, this is shown in the lists by linking the labels with /, as in *crayfish/lobster/prawn* or *moss/seaweed*. Where one proto-gloss has been associated with two or more distinctively different proto-forms, this is indicated in the lists by (1), (2), etc. after the gloss, but forms have not been regarded as distinctly different merely because of variations in individual reconstructive conventions, as in PEO *qupi* and ?uvi *yam* (Wurm and Wilson 1975: 244) or POC *pi(n)so* and *bi(s,z)o reed* (Wurm and Wilson 1975:167).

3.1 POC reconstructed glosses

There are, for POC, 69 flora items and 67 fauna items in the data base. These totals are surprisingly high, given the lack of specific concern with the flora and fauna fields on the part of many of those who compiled the lexicons available for comparison, and of most of the comparativists who produced the reconstructions.

3.1.1 Flora

LEVEL 1: LIFE FORMS

grass
tree
vine/creeper [total 3]

GENERICIS

Alocasia (sp.)⁵/*plant* sp. (*tuberous*)
bamboo (sp.)
banana (sp.)
coconut (?)
Colocasia (sp.)/*taro* (sp.) (1)
cucumber (?)
Curcuma longa (*turmeric*)⁶/*plant* sp. (*with yellow rods*) (?)

ginger (1) (?)⁶
ginger (2) (?)
kava
melon (?)
moss/seaweed (?)
nettle/tree nettle/Laportea (?)
plant sp. (unspecified) (?)
rattan (?)
reed (?)
sago (?)
sugarcane
taro (sp.) (2)
tobacco plant (?)
turmeric (curcuma) (?)⁶
yam
yams (wild) (?) [total 23]

LEVEL 2: GENERICS

[kinds of tree]
areca nut
banyan
Barringtonia
betel pepper
breadfruit (1)
breadfruit (2)
Calophyllum inophyllum
Canaga odorata
candlenut (Aleurites moluccana)
casuarina (sp.)
Cerbera
cinnamon tree
citrus (lemon, orange)
Cordyline (sp.)
Erythrina/shade tree
Ficus (sp.)
Freycinettia banksii
hardwood (1)
hardwood (2)
hibiscus
Malay apple (Syzygium malaccense)
mangrove (Rhizophore)
Morinda citrifolia
mulberry sp. (paper mulberry, Octomeles sumatrana) (1)
mulberry sp. (paper mulberry, Octomeles sumatrana) (2)
pandanus (1)
pandanus (2)
puzzlenut tree
sandalwood/Santalum
Spondias dulcis (mango) (1)
Spondias dulcis (mango) (2)
*Tahitian chestnut (Terminalia)*⁷
teak
*Terminalia (sp.) (Tahitian chestnut)*⁷

tree sp. (coastal) (1)
tree sp. (coastal) (2)
tree sp. (fragrant)
tree sp. (unspecified) (1) through (6) [total 43]

3.1.2 Fauna

LEVEL 1: LIFE FORMS

*animal*⁹
bird
*fish*⁹
snake
*"wug"*¹⁰ [total 5]

GENERICIS

butterfly/moth (?)
cassowary (?)
clam (?)
cowrie (sp.)/mollusk sp. (2) (?)¹¹
crab (sp.) (1) (?)
crab (sp.) (2) (?)
*crayfish/lobster/prawn*¹² (?)
crocodile (?)
mollusk sp. (1) (?)
mother of pearl shell (?)
mussel (1) through (3), all (?)
oyster (?)
oyster sp. (?)
sea urchin (?)
shellfish (sp.) (?)
spider (?)
turtle (?) [total 19]

LEVEL 2: GENERICIS

[*kinds of animal*]¹³
dog
flying fox/bat
pig (1)
pig (2)
rat (1) through (3) [total 7]

 [*kinds of bird*]
bird sp. (unspecified) (1)
bird sp. (unspecified) (2)
*dove*¹⁴
fowl (domestic)
frigate bird
gannet
*pigeon*¹⁴
sea-gull/sea-swallow/Sterna (sp.) [total 8]

[kinds of fish]
bonito
cuttlefish (1)/squid/octopus (2)
cuttlefish (2)
eel (sp.)
fish (scavenging)/scavenger fish/fish sp.
(unspecified) (1)/yellow-finned groper
fish sp. (small)/fish sp. (unspecified) (2)
fish sp. (unspecified) (3) through (5)
mullet
octopus (1)
octopus (3)
porpoise/dolphin
shark (sp.)
stingray [total 15]

[kinds of "wug"]¹⁵
ant (sp.)/cockroach (sp.)/crab (sp.) (3)/worm (2)
balolo worm
caterpillar
centipede
fly (1)
fly (2)/maggot/worm (1)
grub/termite/white ant
housefly
insect (sp.)
louse (body)
louse (head)
mosquito
wood weevil [total 13]

3.2 PEO reconstructed glosses

There are, for PEO, 23 flora items and 32 fauna items in the data base. These totals are markedly lower than those for POC, PML and PPN. In all probability this is merely a reflection of the relatively small total amount of PEO vocabulary that has as yet been reconstructed.

3.2.1 Flora

LEVEL 1: LIFE FORMS

tree
vine [total 2]

GENERICS
banana (sp.)
coconut
Colocasia (sp.)/taro (sp.)
moss/seaweed (?)
mushroom (sp.) (?)
sugarcane
yam [total 7]

LEVEL 2: GENERICS

[kinds of tree]
 areca nut
 Barringtonia
 Canaga odorata
 hibiscus
 mangrove (Rhizophora)
 orange
 pandanus (1) through (3)
 Tahitian chestnut (Terminalia)
 teak
 tree sp. (unspecified)(1)
 tree sp. (unspecified)(2) [total 13]

 [kinds of vine]
 creeper (used for fish poison) [total 1]

3.2.2 Fauna

LEVEL 1: LIFE FORMS

animal⁸
 bird
 fish
 snake
 "wug"¹⁶ [total 5]

GENERICS

butterfly/moth (?)
 clam (?)
 cowrie (sp.)(?)
 crayfish/lobster/prawn (?)
 crocodile (?)
 mussle (sp.)/shellfish sp. (bivalve) (1)(?)
 mussel (sp.) (black, used for bonito hooks and shell money)(?)
 shellfish (sp.)/shellfish sp. (bivalve) (2)(?)
 spider (?)
 turtle (?) [total 10]

LEVEL 2: GENERICS

[kinds of animal]¹³
 pig
 rat [total 2]

 [kinds of bird]
 bird sp. (tropic, Phaet[h]on) [total 1]

 [kinds of fish]
 bonito
 fish (with poisonous spines)
 fish sp. (unspecified)
 mullet
 octopus
 shark (sp.)
 stingray [total 7]

*[kinds of "wug"]*¹⁷
ant (sp.)
balolo worm
caterpillar
fly (1)
fly (2)
*louse (head)*¹⁸
mosquito [total 7]

3.3 PML reconstructed glosses

There are, for PML, 40 flora items and 61 fauna items in the data base. That these totals are relatively high, and that flora and fauna items constitute about 15% of the total set of reconstructions produced for PML (Levy and Smith 1969), presumably reflects some specific interest in the flora and fauna fields on the part of the comparativists.

3.3.1 Flora

LEVEL 1: LIFE FORMS

grass
*tree*¹⁹
vine (3) [total 3]

GENERICIS

bamboo (sp.)
coconut
fern sp. (Lycopodium)
ginger (?)
moss (?)
rattan (?)
reed (?)
sago (?)
sugarcane
taro (sp.)
yam (1)
yam (2)(?) [total 12]

LEVEL 2: GENERICIS

[kinds of grass]
cuscus [total 1]

[kinds of tree]
areca nut
banyan
Barringtonia
betel pepper
breadfruit
Calophyllum inophyllum
Canarium almond
Cordyline (sp.)
croton tree

hardwood
hibiscus
ironwood
mango
Morinda citrifolia
mulberry sp. (*paper mulberry, Octomeles sumatrana*)
pandanus
Polynesian plum
Terminalia sp.
tree sp. (*flowering*)
tree sp. (*used for houseposts*)
tree sp. (*unspecified*) [total 21]

[kinds of vine]
creeper (1)/*vine* (2)
creeper (2)
vine (1) [total 3]

3.3.2 Fauna

LEVEL 1: LIFE FORMS

bird
*fish*²⁰
snake
*"wug"*¹⁶ [total 4]

GENERICIS

butterfly/moth (?)
clam (?)
cone shell/trochus shell (1)(?)
conch (?)
cowrie (sp.)(?)
crayfish (?)
crocodile (?)
dog (?)
dugong (?)
flying fox/bat (?)
gold lip pearl shell (?)
limpet (?)
mother of pearl shell (?)
mussel sp. (*black, used for bonito hooks and shell money*)(?)
nautilus (*chambered*) (?)
nerite (?)
oyster sp. (*giant wing oyster*)(?)
phalanger (*cuscus*) (?)
pig (?)
tree lizard (*Corucia zebrata*) (?)
trochus shell (2)(?)
turtle (?) [total 22]

LEVEL 2: GENERICIS

[kinds of bird]
bird sp. (*believed to bring omens*)

duck
 eagle
 fowl/hen
 frigate bird
 hawk
 hornbill
 kingfisher (large, shore-dwelling)
 osprey
 pigeon
 seabird (sp.)
 starling sp. (shiny starling, *Callornis metallica*)
 swift/kingfisher [total 13]
 [kinds of fish]
 barracuda
 bonito
 fish sp. (poisonous)/toadfish
 fish sp. (walking)/*Periophthalmus*
 lamprey
 shark (sp.)
 squid
 stingray
 swordfish [total 9]
 [kinds of "wug"]¹⁷
 ant (sp.)
 ant sp. (yellow, with painful bite)
 beetle sp. (bores yams)
 caterpillar
 cricket (sp.)
 dragonfly
 firefly
 fly
 grasshopper
 sandfly
 scorpion
 termite
 worm [total 13]

3.4 PPN reconstructed glosses

There are, for PPN, 79 flora items and 147 fauna items in the data base. These high figures reflect both the sheer quantity of PPN reconstructions available in the data base sources, especially Biggs, Walsh and Waqa 1970, and the fact that several recent Polynesian comparativists have had some specific interest in the flora and fauna fields.

3.4.1 Flora

LEVEL 1: LIFE FORMS

grass
 tree²¹
 vine/creeper [total 3]

GENERICIS

Alocasia (sp.)
arrowroot (?)
bamboo (sp.) (1)
banana (sp.) (2)
banana (sp.)
coconut (1)
coconut (2)
Colocasia sp./*taro* (sp.)
Curcuma domestica (?)
Fagraea berteriana (or other plant with showy leaves) (?)
fern (sp.) (1) (?)
fern (sp.) (2) (?)
fern (sp.) (4) (?)
gardenia (?)
kava
Leptocarpus simplex (?)
moss/seaweed (?)
nettle/tree nettle/Laportea (?)
oxalis sp. (?)
plant sp. (scented) (?)
plant sp. (unspecified) (1) through (6), all (?)
plantain (?)
reed (1) (?)
reed (2) (?)
sedge sp./*fern* (sp.) (3) (?)
sweet potato sp.
turmeric (?)
yam [total 33]

LEVEL 2: GENERICIS

[kinds of tree]
Barringtonia/tree sp. (coastal) (1)
breadfruit (1) through (3)
Canaga odorata
casuarina/tree sp. (unspecified) (22)
Cerbera/tree sp. (unspecified) (8)
citrus
Colubrina asiatica
Cordyline (sp.)
Ficus (sp.)
hibiscus
mango/Spondias dulcis
mangrove
Morinda citrifolia/tree sp. (unspecified) (12)
mulberry sp. (paper mulberry, *Octomeles sumatrana*)
palm (sp.)
pandanus (1)
pandanus (2)
Parinarium insularum/tree sp. (unspecified) (18)
Pritchardia pacifica
puzzlenut tree
sandalwood (*Santalum*)/*tree* sp. (unspecified) (23)
Terminalis (sp.) (*Tahitian chestnut*)/*tree* sp. (coastal) (2)

Thespesia populnea/tree sp. (unspecified) (10)
tree sp. (unspecified) (1) through (7)
tree sp. (unspecified) (9)
tree sp. (unspecified) (11)
tree sp. (unspecified) (13) through (17)
tree sp. (unspecified) (19) through (21)
tree sp. (unspecified) (24) [total 43]

3.4.2 Fauna

LEVEL 1: LIFE FORMS

*animal*⁸
bird^{2 2}
fish^{2 3}
snake
"wug"^{1 6} [total 5]

GENERICIS

bat (?)
butterfly/moth (1) (?)
clam (?)
cowrie (sp.) (?)
crab (sp.) (1) (?)
crab (sp.) (2) (?) / *shellfish (sp.)* (2) (?)
crab (sp.) (3) (?)
crab (sp.) (4) (?)
crayfish (?)
echinoderm/sea egg (?)
hermit crab (?)
Holothuria/sea slug (2) (?)
land crab (1) (?)
land crab (2) (?)
lizard (sp.) (?)
marine animal (unspecified) (1) through (3), all (?)
marine animal (unspecified) (5) (?)
mollusk sp./shellfish (sp.) (7) (?)
moth (?)
mussel (sp.) (?)
oyster (?)
sea slug (1) / *slug/snail* (?)
sea urchin (?)
shellfish (sp.) (1) (?)
shellfish (sp.) (3) through (6), all (?)
spider (1) (?)
spider (2) (?)
sponge (?)
turtle/tortoise (?) [total 34]

LEVEL 2: GENERICIS

[kinds of animal]^{1 3}
dog
pig
rat [total 3]

[kinds of bird]

bird sp. (*tropic, Phaet[h]on*)/bird sp. (*unspecified*) (13)bird sp. (*unspecified*) (1) through (8)bird sp. (*unspecified*) (10) through (12)

fowl

frigate bird (1)

frigate bird (2)

gannet (1)

gannet (2)

hen bird

heron/bird sp. (*unspecified*) (9)

parakeet sp.

parrot sp.

pigeon (1)

pigeon (2)

seabird (sp.) (1)

seabird (sp.) (2) [total 25]

[kinds of fish]

barracuda

bonito (1)

bonito (2)

cavally/fish sp. (*unspecified*) (29)cuttlefish/squid/marine animal (*unspecified*) (4)

dolphin

eel (sp.) (1) through (3)

fish sp. (*unspecified*) (1) through (22)fish sp. (*unspecified*) (24) through (27)fish sp. (*unspecified*) (31) through (38)fish sp. (*unspecified*) (41)fish sp. (*unspecified*) (42)flounder/fish sp. (*unspecified*) (39)

flying fish

garfish/fish sp. (*unspecified*) (40)/swordfish (1)leatherjacket/fish sp. (*unspecified*) (30)

mullet

octopus

shark (sp.) (1) through (3)

stingray

surgeon fish/fish sp. (*unspecified*) (23)

swordfish (2)

Trachurops/fish sp. (*unspecified*) (28)

whale (sp.) [total 59]

[kinds of "wug"]¹⁷

ant (sp.)

ant sp. (*big*)

balolo worm

beetle sp.

caterpillar (1)

caterpillar (2)/maggot (2)/worm (2)

cockroach (sp.)/cricket (sp.) (2)

cricket (sp.) (1)

flea

fly

grasshopper

grub (1)/*threadworm*
grub (2)
housefly
insect sp.
louse (head)
maggot (1)
mantis
mosquito
termite
worm (1) [total 21]

4. DISTRIBUTION OF ITEM TYPES THROUGH HIERARCHIC LEVELS — SUMMARY AND COMMENT

The distribution of item types through hierarchic levels is summarised below:

LEVEL	ITEM TYPE	POC		PEO		PML		PPN		TOTAL
		flora	fauna	flora	fauna	flora	fauna	flora	fauna	
0	unique beginner	-	-	-	-	-	-	-	-	-
1	life form generic	3	5	2	5	3	4	3	5	30
2	generic specific	23	19	7	10	12	22	33	34	160
2	generic specific	43	43	14	17	25	35	43	108	328
3	specific varietal	-	-	-	-	-	-	-	-	-
4	varietal	-	-	-	-	-	-	-	-	-

4.1 Unique beginners

The lack of *unique beginners* in POC, PEO, PML and PPN reflects their non-presence as labelled categories in the Oceanic languages, an aspect of these languages which is consistent with Berlin, Breedlove and Raven's finding that "In folk taxonomies it is quite common that the taxon found as a member of the category *unique beginner* is not labelled linguistically by a single habitual expression." (1974:26).

4.2 Life forms

The support for a given reconstructed *life form* in terms of the number of reconstructed *generics* subsumed by it is as follows:

	POC	PEO	PML	PPN
FLORA				
<i>grass</i>	-	-	1	-
<i>tree</i>	43	13	21	43
<i>vine</i>	-	1	3	-
FAUNA				
<i>animal</i>	7 ²⁴	2	-	3
<i>bird</i>	8	1	13	25
<i>fish</i>	15	7	9	59
<i>snake</i>	-	-	-	-
"wug"	13	7	13	21

These figures show that only *tree* in the flora field, and *bird* (PEO excepted), *fish* and "wug" in the fauna field, have their life form status supported by clear evidence²⁵ of generic polytypicism. The evidence is less satisfactory for *animal*, perhaps because of the small range of mammalia in the Oceanic region. The other items, *grass*, *vine*, and *snake*, must here be regarded as having only notional *life form* status, based on their association with major observable discontinuities within the flora or the fauna field, and, in the cases of *grass* and *vine*, also on their considerable potential for generic polytypicism.

That the *life form* status of "wug" is supported by clear evidence of generic polytypicism is some vindication of the setting up of this category for POC on the basis of overlapping proto-glosses (ref. note 10) and for PEO, PML and PPN by analogy with the POC situation. There may well turn out to be similar grounds for proposing *life forms* of comparable status to "wug" with glosses such as "shellfish", "crustacean", "bush/shrub" and "grerb" (for which latter see Brown 1984:13).

The specific composition (*tree* + *grass* + *vine*) of the flora *life form* sets for POC, PML and PPN is consistent with stage four of Brown's botanical *life form* encoding sequence (1984:24). The PEO set (*tree* + *vine*) is a relatively rare combination (ref. Brown 1984:25) which may well merely reflect the comparatively small stock of reconstructions as yet available for this proto-language.

The specific composition (*bird* + *fish* + *snake* + "wug" + *animal*) of the fauna *life form* sets for POC, PEO and PPN is consistent with stage five of Brown's zoological *life form* encoding sequence (1984:24), and the composition (*bird* + *fish* + *snake* + "wug") of the PML set is consistent with stage four. Given the reservations which must be entertained concerning the viability of *animal* [mammal] as a *life form* item in the Oceanic context, the composition of the fauna *life form* sets of all four proto-languages may well in fact be consistent with stage four of Brown's sequence.

4.3 Generics

Of the 518 items listed in Section 3, 488 are *generics* — 160 at level 1 and 328 at level 2. These figures are consistent with Berlin, Breedlove and Raven's claim that "In typical folk taxonomies, the taxa that are members of

the ethnobiological category, *generic*, are much more numerous than life form taxa, but are nonetheless finite, ranging in the neighborhood of 500 classes." (1974:26).

The 160 level 1 *generics* include a high proportion of items followed by (?), many of which may well turn out to be level 2 *generics*, each subsumed under some ('notional') *life form* item glossable as "*shellfish*", "*crustacean*", "*bush/shrub*", "*grerb*", etc. (ref. Section 4.2). Some other level 1 *generics* followed by (?) may well turn out to be *bona fide* level 1 items because of cultural salience as totems, e.g. *turtle*, *clam*; or because of anomalous qualities, e.g. *cassowary* (ref. Bulmer 1967), *moss/seaweed*, *spider*.

The level 1 *generics* which are not followed by (?) include staple foods, e.g. *yam*, *taro*, *coconut*; and items of cultural salience, e.g. *kava*, *bamboo*. All of these items are associated with extensive *specific*, and often also *varietal*, subclassification in many Oceanic languages.

Every level 2 *generic* is subsumed by a *life form*. If a given *life form*, such as *animal* or "*wug*", turned out not to be viable, then the *generics* subsumed by it would have to be re-located as level 1 items.

4.4 Specifics and varieties

That there are no *specifics* or *varietals* among the reconstructed glosses is consistent with the tendency for such items to be multi-morphemic, with glosses such as *red-leaved banyan* and *white-leaved banyan*, and often also to be metaphoric (ref. Walsh 1980). Complex items of this kind are less likely to share the patterned form-plus-meaning resemblances between languages which make reconstruction possible than are simple items, and the likelihood of such resemblances for complex items decreases rapidly as the range of related languages involved in a given comparison increases. Therefore, *specifics* and *varietals* are much more likely to be reconstructable for lowest-order subgroups than for any higher-order subgroups, and the likelihood of their reconstructability decreases rapidly as the subgrouping hierarchy is ascended.

5. CONCLUSION

The expectation stated in the last paragraph of *Section 2* has been fulfilled to a limited extent. For POC the reconstructions include not only higher order items — *life forms* and level 1 *generics*, many of which are culturally salient — but also, contrary to expectation, a relatively large number (86) of level 2 *generics*. For the sub-Oceanic proto-languages the reconstructions include not only virtually all the *life forms* present for POC, together with similar ranges of level 1 *generics*, but also, in the case of PPN only, a markedly greater number (108 as against 43 for POC) of level 2 *generics* in the fauna field.

However, the kind of operation that has been reported on in this paper is grossly constrained by the essentially etic quality of the reconstructed data (ref. *Section 1*). There may well be in present-day Oceanic languages, and there may well have been in earlier stages of these language traditions, broadly inclusive classes with boundaries approximate to those implied, in etic terms, by such *life form* glosses and proto-glosses as, e.g., *fish*, *bird*,

or *tree*, but we cannot, from etic data, know whether, e.g., POC *fish* includes *octopus*, *shark*, or *stingray*. If, in emic terms, any of these latter items is not included either under what is glossed as *fish* or under some other *life form*, then such an item or items would have to be listed among the level 1 *generics* rather than the level 2 *generics*.

Until this kind of emic data is available it is not possible to establish at all precisely what the detail of a reconstructed hierarchy for flora or fauna should be. What is required, then, for the Oceanic languages, is the kind of emic lexical and hierarchic data in the flora and fauna fields that are available for Tzeltal in Berlin, Breedlove and Raven 1974 and Hunn 1977. With such data it would be possible to reconstruct, for POC and for sub-Oceanic proto-languages, taxonomic hierarchies for flora and fauna that were at least notionally emic. It would then also be possible, *inter alia*, to carry out at a much higher level of significance the kind of operation that has here been reported on.

NOTES

1. This paper is a later version of "The lower-order reconstruction of flora and fauna nomenclature within EAN" which was presented at the 15th Congress of the Pacific Science Association. Funding for conference travel and underlying research was provided by the University of Sydney.
2. Due allowance has been made for the fact that Wurm and Wilson (ref. 1975: viii-xiv) to some extent 'processed' glosses for convenience of presentation.
3. In most cases the relevant flora and fauna data were not the result of any systematic and comprehensive investigation of these lexical fields.
4. "Covert" here does not have the connotation of "intermediate" ascribed to it by Berlin, Breedlove and Raven (1974:27). In the present context "notional" might be a more appropriate term.
5. The use of (*sp.*) in contexts such as this follows Wurm and Wilson — "Where a reconstruction may designate either a species or a generic term, and it is not necessary to draw the distinction, the gloss assumes the following form: *BAMBOO (SPECIES)*." (1975:xii).
6. There is some imprecision in the use of *Curcuma longa*, *turmeric* and *ginger* in the underlying data.
7. Each of these glosses — *Tahitian chestnut (Terminalia)* and *Terminalia (sp.) (Tahitian chestnut)* — is associated with a different POC form, ref. Wurm and Wilson 1975:12 and 15.
8. The gloss *animal* is presumed to signify *mammal*. There are, in the underlying data, some grounds for doubt that *animal [mammal]* can validly be reconstructed as a distinct category for POC, PEO or PPN.
9. As general terms for *fish*, POC manu(k) and tuRi(a) (Wurm and Wilson 1975: 77) are here disregarded.

10. The term "*wug*" has been borrowed from Brown (1984:16). This category is here proposed as a *life form* on the basis of the semantically partially overlapping POC kalo *ant, cockroach, crab, worm* (Grace 1969:50) and POC qulo(s) *worm, fly, maggot, (prawn ?)* (Grace 1969:82). POC mpaya and ŋmata *worm* (Wurm and Wilson 1975:243) are here disregarded.
11. These two glosses are equated on the basis of POC mpule (Wurm and Wilson 1975:46 and 131). The other gloss for this form — *shell, white* (Wurm and Wilson 1975:186) — is here disregarded.
12. POC qada(ŋ) *prawn* (Wurm and Wilson 1975:158) is a misprint for POC quda(ŋ) (Grace 1969:82).
13. If *animal* is not regarded as a viable *life form* item for POC, PEO and PPN, then the *generics* here subsumed by it would have to be re-located as level 1 items.
14. Wurm and Wilson (1975:152) subsume POC dupe and mpune under *pigeon, dove*, whereas Grace has POC dupe *dove* (1969:49) and POC mpune *pigeon* (1969:68).
15. If "*wug*" is not regarded as a viable *life form* item for POC, then the *generics* here subsumed by it would have to be re-located as level 1 items.
16. PEO, PML and PPN "*wug*" must be regarded as covert or notional *life forms*, here proposed by analogy from POC "*wug*".
17. If "*wug*" is not regarded as a valid covert or notional *life form* for PEO, PML and PPN, then the *generics* here subsumed by it would have to be re-located as level 1 items.
18. PEO li(s,z)a *louse (unspecified)* (Wurm and Wilson 1975:125) probably should be glossed *louse egg* and is therefore disregarded.
19. PML o'a *tree* (Wurm and Wilson 1975:224) is glossed *tree, cavity in tree* in Levy and Smith (1969:17) and is therefore disregarded.
20. PML roma *fish* (Wurm and Wilson 1975:77) is glossed *poisonous substance in shellfish, crabs, fish; poison fish with Barringtonia* in Levy and Smith (1969:18) and is therefore disregarded.
21. No explanation is here offered for PPN fuqu, raqakau, ra'a and kau *tree* (Wurm and Wilson 1975:224) — perhaps the specific sources of the reconstructions will provide it.
22. As a general term for *bird*, PPN moso (Wurm and Wilson 1975:18) is here disregarded.
23. As a general term for *fish*, PPN ma-lau (Wurm and Wilson 1975:77) is here disregarded.
24. This figure is inflated by the reconstruction of two terms for *pig* (Wurm and Wilson 1975:152) and three terms for *rat* (Wurm and Wilson 1975:165).
25. i.e. by having three or more *generics* subsumed by a given *life form* in POC, PEO, PML and PPN.

BIBLIOGRAPHY

- BERLIN, B., D.E. BREEDLOVE and P.H. RAVEN
 1974 *Principles of Tzeltal plant classification*. New York and London: Academic Press.
- BIGGS, B., D.S. WALSH and J. WAQA
 1970 *Proto-Polynesian reconstructions with English to Proto-Polynesian finder list. Working Papers in Linguistics*. Auckland: Department of Anthropology, University of Auckland.
- BROWN, C.H.
 1981 Growth and development of folk zoological life-forms in Polynesian languages. *Journal of the Polynesian Society* 90:83-110.
 1982 Growth and development of folk botanical life-forms in Polynesian languages. *Journal of the Polynesian Society* 91:213-243.
 1984 *Language and living things: uniformities in folk classification and naming*. New Brunswick, New Jersey: Rutgers University Press.
- BULMER, R.N.H.
 1967 Why is the cassowary not a bird? A problem of zoological taxonomy among the Karam of the New Guinea Highlands. *Man*, n.s. 2:5-25.
- HUNN, E.S.
 1977 *Tzeltal folk zoology: the classification of discontinuities in nature*. New York, San Francisco, London: Academic Press.
- GRACE, G.W.
 1969 A Proto-Oceanic finder list. *Working Papers in Linguistics* 2:39-84. Honolulu: Department of Linguistics, University of Hawaii.
- LEVY, R. and N. SMITH
 1969 *A Proto-Malaitan lexicon*. Department of Anthropology, University of California at Santa Cruz. Mimeographed.
- PUKUI, M.K. and S.H. ELBERT
 1965 *Hawaiian dictionary: Hawaiian-English, English-Hawaiian*. 3rd edition. Honolulu: University of Hawaii Press.
- WALSH, D.S.
 1980 Patterns of metaphor in Raga plant nomenclature. *Ethnomedizin* 6/1-4:127-137.
- WALSH, D.S. and B. BIGGS
 1966 *Proto-Polynesian word list I*. Auckland: Linguistic Society of New Zealand.

WILLIAMS, H.W.

1971 *A dictionary of the Maori language*. 7th edition. Wellington: Government Printer.

WURM, S.A. and B. WILSON

1975 *English finderlist of reconstructions in Austronesian languages (post-Brandstetter)*. PL, C-33.

THE POSITION OF ATAYAL IN THE AUSTRONESIAN FAMILY

Paul Jen-kuei Li

1. INTRODUCTION¹

Atayal has hitherto been regarded as one of the most aberrant Austronesian languages, mainly because of its low percentages of cognates with other Austronesian languages; see Dyen (1965a, 1971a). Obviously its poorly understood phonology and morphology has constituted part of the difficulty in identifying a number of cognates. Recent detailed accounts of Atayal phonology (including phonetics and morphophonemics) given in Li (1980a) and the differences between men's and women's speech in Atayal (Li 1980b, 1982c) and the reconstruction of Proto-Atayalic phonology (Li 1981) should provide a more solid basis for determining cognates.

Another reason for the seemingly aberrant nature of Atayal is that most of the previous studies on the language, such as Ogawa (1931), Egerod (1965a,b, 1966a,b, 1980), Yamada and Liao (1974), and Tsuchida (1976), were based on Sqliq, the most innovative dialect in the entire Atayalic group.

In this paper I shall examine my field data for Mayrinax, the most conservative dialect of Atayal, identify the Austronesian cognates, and compare Mayrinax with the other Formosan languages. I shall try to argue that Atayal and Sediq are not an isolated subgroup of Formosan languages as Dyen and Ferrell (1969) have assumed. On the contrary, they have closer relationships with some other Formosan languages, i.e. Saisiat, Pazeh, Taokas, Babuza, Papora and Hoanya, as based on both phonological and lexical evidence.

2. VARIANT LEXICAL FORMS IN THE ATAYALIC DIALECTS

It is difficult to reconcile the differences of the lexical forms in various Atayalic dialects² without consulting material on Mayrinax, which still preserves some distinctions in male and female forms of speech. For example,

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 257-280. *Pacific Linguistics*, C-88, 1985.

© Paul Jen-kuei Li

		<i>anteater</i>	<i>maple tree</i>	<i>winnowing basket</i>	<i>brains, marrow</i>
Ata	Sq	qom	raga?	luku?	luqus
	Ms	?agum	raa?	baluku?	lu?us
	Sk	q-qom	raa?	bluku?	luqi?
	Mx	qaum/qagum ³	raa?/raga?	baluhiŋ/baluku?	luqi?/luqus
	Pl	?oŋ	raa?	baluku?	--
Sed	Tn	?aruŋ	dara?	--	luqi?
	Td	?aruŋ	dara?	bluhiŋ	luqi?
	In	?aruŋ	dara?	bluhiŋ	luqi?
PA		*qaum/qagum	*daa?/daga?	*baluhiŋ/baluku?	*luqi?/luqus
		<i>clean (water)</i>	<i>dig</i>	<i>drunk</i>	<i>gray hair</i>
Ata	Sq	m-tasaw	k-m-e-huy	busuk	quri?
	Ms	ma-tasi?	k-am-ai-huy	ma-bas-in-uk	?uri?
	Sk	m-tasaw	k-m-e-huy	m-s-in-uk	?i-quri?
	Mx	ma-tasiq/ ma-tasaw	k-um-ai-huw/ k-um-ai?	bus-in-uk/ ma-busuk	quri?/quras
	Pl	m-tasi?	k-um-e-hur	bas-n-uk-an	?uri?
Sed	Tn	m-taso	k-m-ari?	bsuk-an	qudi?
	Td	m-tasaw	k-m-ari?	bsuk-an	qudas
	In	m-tasag	k-m-ari?	bsuk-an	quji?
PA		*ma-tasiq/ ma-tasaw	*k-um-ari-huw/ k-um-ari?	*bus-in-uk/ ma-busuk	*qudi?/qudas

3. ADDITION OF AFFIXES TO THE AUSTRONESIAN COGNATES IN ATAYAL

Due to the innovations in morphological shapes, some Atayalic cognates are not readily recognisable, e.g. PAN *bulat > Sq bya-ciŋ, Ms baya-tiŋ, Sk byal-iŋ, Mx bua-tiŋ, Mt buya-tiŋ, Pl bura-tiŋ *moon*; cf. PAN *kuCuh > Sq, Ms, Sk, Mt, Pl ku-hiŋ *head louse*. Thus, the percentage of cognates between Atayal and other Austronesian languages must have been deflated to a certain extent.

A comparison between the male and female forms of speech in Mayrinax can help account for many unusual derivations, i.e. the morphological structure in the Atayal and Sediq languages. It is important to note that the affixes originally added to the male forms now appear in a number of lexical items used by both sexes in various Atayalic dialects. The corresponding forms reconstructed for PAN lack these affixes. The following are some examples of this phenomenon. (PAN stands for Proto-Austronesian and PA for Proto-Atayalic.)

1. (a) The infix -a-, -ya- < PA *-ra-:
PAN *qauy > Mx qau-a-g, Mt ?au-ya-g *type of bamboo*
- (b) The infix -ra-, -ya- < PA *-ra- or -na- < PA *-na-:
PAN *(dD)aqiS > Sq rqi-ya-s, Ms, Mt ra?i-ya-s; Sk rqi-na-s,
Mx raqi-na-s, Tn dqe-ra-s, Td, In dqe-ra-s *face*
PAN *Səpat > Sq pa-ya-t, Ms sapa-ya-t, Pl pa-ra-c *four*
2. The infix -in- or -n- < PA *-in-:
PAN *butuy > Ms bah-in-uw, Mx buh-in-ug *bow*
PFN⁴ *DakəS > Sq k-n-us, Ms, Pl rak-in-us, Sk rk-n-as *camphor laurel*

3. The suffix *-li?*, *-lit*, *-lic* < PA **-lid*
 PAN **laŋaw* > Sq, Sk *ŋ-li?*, Pl *raŋ-lic* *fly (insect)*
 PAN **qabuŋ* > Sq, Sk *qbu-li?*, Mx *qabu-li?*, Pl *ʔabu-lic*, Sed *qbu-lic* *ashes*
4. (a) The suffix *-niq*, *-ni?* or *-naq* < *PA *-niq*:
 PAN **butuŋ* > Pn *bhu-niq*, Mt, Pl *bahu-ni?*, Tn *bh-e-naq*, Td,
 In *bh-ə-niq* *bow*
- (b) The suffix *-iq*:
 PAN **ʔuŋaC* > Mx *ʔuw-iq*, Sq, Sk *ʔuŋ-iq*, Bl *ʔuŋ-iq* *vein*
5. The suffix *-qiy*, *-ʔiy* or *-ʔig* < PA **-qig*:
 PAN **tuŋa* > Sq, Sk *tla-qiy*, Ms *tala-ʔiy*, Mx *tula-qiy*, Mt *tula-ʔig* *eel*
6. The suffix *-ti?* or *-ci* < PA **-ti?*:
 PA **baŋaŋ* > Ms *bagaci?*, Mx *bagati?* *Alocasia*
7. The suffix *-nux*:
 PAN **Sagək* > Pl *s-uŋ-k-a-nux*, Tn *s-m-k-e-nux*, Td *s-m-k-ə-nux*,
 In *p-sk-ə-nux* *to smell*
 PAN **butuŋ* > Lm *bh-i-nux*, Sk *bh-e-nux*, Mn *bh-ai-nux*, Mt *bah-i-nux* *bow*
- cf. 2, 4(a) above. Mt. has *bahu-ni?* ~ *bah-i-nux*, as based on different informants.
- PAN **batu* > Sq *tu-nux*, Sk *btu-nux*, Mx, Pl *batu-nux* *stone*
 PAN **[ʔ]ipən* > Sq *ʔ-nux*, Lm *gʔ-nux*, Ms, Mt, Sx *gaʔa-nux*, Mx *giʔ-nux/gipun*,
 Tl *ʔa-nux*, Bl *gʔ-nux*, Pl *ʔapa-nux* *tooth*
8. The suffix *-al*:
 PAN **kiŋa* > Mx *k-um-ita-al* *to see*
 PAN **Səpi* > Ms *s-m-piy-al*, Mx, Mt *s-um-apy-al*, Sx *s-am-py-al*,
 Tl *ma-sapiy-al*, Pl *ma-spel* < **ma-səpi-al* *to dream*
9. The suffix *-iŋ* or *-Ciŋ* (where C stands for a consonant):
 PAN **tuDuq* > Mx *m-si-tur-iŋ*, Sk *s-tur-iŋ*, Ms *ma-s-tur-iŋ*, Sx *ma-s-turiŋ*,
 Tl *sa-tur-iŋ*, Sed *tudiq* *to drip*
 PAN **bulaŋ* > Sq *bya-ciŋ*, Sk *byal-iŋ*, Pl *bura-tiŋ*, Mx *bua-tiŋ*,
 Bl *bya-ciŋ* *moon*

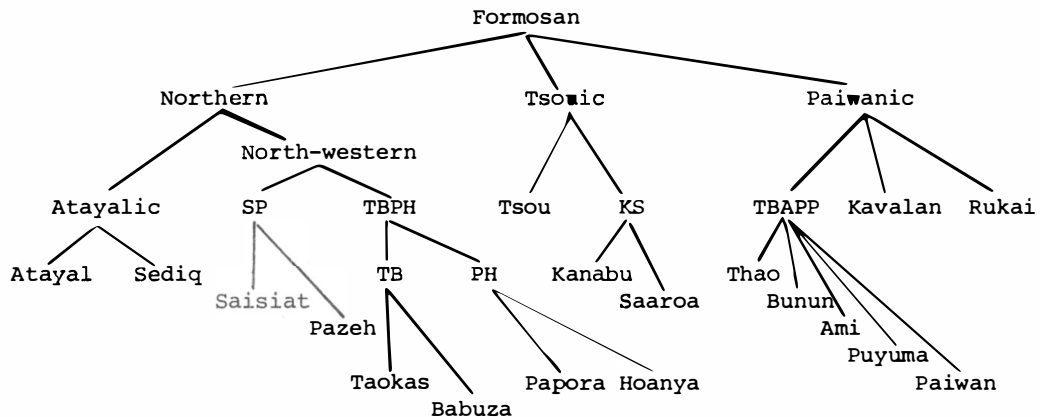
See Li (1980b:14-15) for other types of affixes and examples.

4. THE HYPOTHESIS OF A NORTHERN SUBGROUP OF LANGUAGES

The classification of the Formosan languages has been the subject of discussion for many years, yet the issue is far from settled. Both Dyen (1965a, 1971b) and Ferrell (1969) made a tripartite classification of the Formosan languages into Atayalic, Tsouic and Paiwanic. The Atayalic group is comprised of Atayal and Sediq, the Tsouic group of Tsou, Kanakanabu and Saaroa, and the Paiwanic group of the rest. Ferrell (1969:25) further divided the Paiwanic into two subgroups: Paiwanic I included Rukai, Pazeh, Saisiat, Luilang, Favorlang, Taokas, Papora, Hoanya, Thao, Paiwan and Puyuma, while Paiwanic II included Bunun, Siraya, Ami, Kavalan/Ketagalan and Yami, on the basis of retention or loss of the phonemic distinction between PAN **t* and **C*. These two investigators based their conclusions mainly on lexicostatistical evidence, using short wordlists for a number of languages. Tsuchida (1976:15) generally followed their classification except that he grouped Rukai with the Tsouic group, on the basis of certain phonological as well as lexical evidence.

In this paper I shall put forth arguments for the hypothesis⁵ of a Northern subgroup of Formosan languages, which is comprised of Atayal, Sediq, Saisiyat, Pazeh, Taokas, Babuza, Papora and Hoanya, on the basis of phonological, lexical and syntactic evidence. Since all these languages are or were located in the north or north-west coast of Formosa, they can be labelled the 'Northern' group. I shall present some phonological evidence showing parallel sound changes in the group in this section, phonological evidence for the internal relationships of the group in Section 5, and lexical evidence for the group in Section 6. The revised classification of Formosan languages is given in Table 1.

Table 1: A tentative classification of Formosan languages



There is some syntactic evidence for a close relationship between the Atayalic group, Saisiat and Pazeh. These languages generally do not distinguish the feature [± personal] in their construction markers; see Ogawa and Asai (1935:26) and Egerod (1965a, 1966a) for Atayal, Starosta (1974:334) for Sediq, Li (1978:600) for Saisiat⁶ and Li (1978:573) for Pazeh. However, such a distinction is clearly made in four sets of construction markers (nominative, accusative, genitive and locative) in the Paiwanic languages such as Paiwan (Ogawa and Asai 1935:137), Rukai (Li 1973:86, Starosta 1974:319), Kavalan (Li 1978:583), Ami (Starosta 1974:301, Chen 1985:127) and Puyuma (Ogawa and Asai 1935:303). This may be regarded as an important piece of syntactic evidence for not grouping Saisiat and Pazeh with the Paiwanic, but rather with the Atayalic.

In absence of syntactic data for the four north-western languages (Taokas, Babuza, Papora and Hoanya), and the inadequate syntactic data for the rest of the languages, phonological and lexical evidence is the most feasible means for classification.

All languages in the Northern group have the same reflexes for the following PAN phonemes: (1) PAN *ʃ (Dahl's 1981 *S₂) > h, (2) PAN *h (Tsuchida and Dahl's *H₁) > h (3) PAN *ń > l, (4) PAN *ʈ (Dyen's *N) > l, (5) PAN *C > c [ts] or s; see Table 2. That is to say, PAN *ʃ and *h have merged as h and PAN *ń and *ʈ have merged as l in all members of the group. The former merger is unique to the Northern group, while the latter merger is found also in some other Formosan languages such as Rukai and Bunun. Furthermore, all languages in the Northern group retain the phonemic distinctions between PAN *t and *C, PAN *n and ʈ, typical Formosan features which have been lost in some Paiwanic languages such as Bunun, Kavalan and Ami.

In addition, PAN *D, *d' and *z have merged as Atayal r, Sediq d, Saisiat r, Pazeh d (devoiced word-finally), Taokas, Babuza and Papora d or t. That is to say, this phonemic merger is found in all languages in the Northern group. However, the same merger is also found in the Tsouic, Bunun, Thao, Kavalan and Ami.

Table 2: Formosan reflexes of Proto-Austronesian⁷

PAN	*p	*b	*t, *T /_i	*C	*d	*D	*d', *Z
Ata Sq	p -p	b -p	t c -t	s	r -?	r -?	r
Sk	p -p	b -p	t c -c	c	r -?	r -?	r
Mx	p -p	b -b	t t -t	c	r -?	r -?	r
Mt	p -p	b -p	t t -t	s	r -t	r -t	r
Pl	p -k	b -k	t t -c	c	ř -c	ř -c	ř
Sed Tn	p -k	b -k	t t -c	c	d -c	d -c	d
Td	p -k	b -k	t t -c	c	d -c	d -c	d
In	p -k	b -k	t c -c	s	d -c	d -c	d
Tsou Dh	p	f	t	c	c	c	c
Kan	p	v	t	c	c	c	c
Sar	p	v	t	c	c	s	s
Ruk Bu	p	b, v/_g'	t	c	d- -D	D	d
Mg	p	b	t	c	d- -D	D	d
Mn	p	v	t	c	ǒ ǒ	ǒ	ǒ
Bun Tk	p	b	t	t	d -?	d -?	d
Is	p	b	t	t		d	d
Pai Bu	p	v	ǃ	c	ǃ -z	D, z	ǃ
Puy Pn	p	b	t	T	d	D	d
Kl	p	v	t	T	ǒ	z	ǒ
Thao	p	f	t	θ	s	s	s
Sai Ta	p	b[β]	t	s	r	r	?, r
Paz	p	b -p	t	s	d -t	d -t	?, d
Kav	p	β	t	t	z	z	z
Ami Sa	p	b	t	t	ř	ǒ	ǒ
Fr	p	v	t	t	ř	ř	L
Tao	p, w	b, v	t	s	t	d, r, t	d, t
Bab	p, Ø	b	t	c, s, ch	t, z	d, r, t	t
Pap	p	b	t	ts, θ, s	r, l	d -Ø	d
Hoa	p	b	t	s		z, s, r, l	dz
Sir	p	v	t	t		s	d

Table 2: Formosan reflexes of Proto-Austronesian (cont'd)

Pan	*k /_q, h	*g	*g' (*j) /i_a	*Y (*R) /a_a/else/i_# /a_u_#	*q
Ata Sq	k q		-r-, -r-, -y	g -g- Ø -y -w	q
Sk	k q		-s-, -s-, -s	g -g- Ø -y -x	q
Mx	k k		-s-, -s-, -y	g -g- Ø -y -g	q
Mt	k k		-s-, -s-, -g	g -g- Ø -g -g	?
Pl	k k		-ř-, -ř-, -y	g -g- Ø -y -w	Ø
Sed Tn	k q		-g-, -y-, -y	g -r- -g- -y -w, -Ø	q
Td	k q		-w-, -y-, -?	w -r- -w- -? -?, -w	q
In	k q		-g-, -y, -g	g -r- -g- -g -g	q
Tsou	?	k, ?	Ø	r	Ø
Kan	k	k	l, Ø/i_, i/i_#, c	r	?
Sar	k	k	ɬ, Ø/i_ c	r	?
Ruk Bu	k	g	g, Ø/i_	r Ø	Ø
Mg	k	g	g, Ø/i_, k/b_	r Ø	Ø
Mn	k	h	h, Ø/i_, k/_i	r -?-	Ø
Bun Tk	k		Ø	l	q
Is	k		Ø	l	χ
Pai Bu	k	g	d	Ø	q
Puy Pn	k	g	d	r	?
Kl	k	h	ð	r	ɸ
Thao	k	h(?)	ð	ɬ, Ø	q
Sar Ta	k		z	L	?
Paz	k	g, k	z	x	?
Kav	{k/i q/_u, ə, a}		n	r/i_i, R, l/	Ø, -?
Ami Sa	k		n, d	L	?, -ɸ
Fr	k		n	L	?, -ɸ
Tao	Ø		t	x, h	h
Bab	Ø		d	g, r	0, h, ch
Pap	Ø		d	d, l, r	Ø
Hoa	Ø		z, dz	h	Ø
Sir	k		n	g	Ø

Table 2: Formosan reflexes of Proto-Austronesian (*cont'd*)

PAN	*t' (*s)	*θ	*s (*S ₁)	*ʃ (*S ₂)	*h (*H ₁)	*h (*H ₂)
Ata Sq	h	h -s	s	h	h	?
Sk	h	h	s	h	h	?
Mx	h	h, x	s	h	h	?
Mt	h	h	s	h	h	?
Pl	h	h	s	h	h	?
Sed Tn	h	h	s	h	h	?
Td	h	h	s	h	h	?
In	h	h	s	h	h	?
Tsou	s	s	s	∅	∅	∅
Kan	∅	s	s	∅	∅	∅
Sar	∅	s	∅	∅	∅	?
Ruk Bu	∅	θ	s		∅	∅
Mg	s	θ	s	∅	∅	∅
Mn	∅	s	?	∅	∅	?-h
Bun Tk	s	c	s	s, ∅	h	?
Is	s	s	s	s, ∅	∅	?
Pai Bu	t	t	s	s	∅	?
Puy Pn	s	s, ∅	∅	∅	∅	?
Kl	s	s	∅	∅	∅	∅
Thao	t	t	ʃ	∅	∅	?
Sai Ta	h	h	ʃ	h	h, ∅	-h, ∅
Paz	z-z-t	z	s	h	h	h-?
Kav	s	s	s	s	-∅-	
Ami Sa	c	c	s	s	h	?
Fr	c	c	s	s	h	?
Tao	t -∅				h	
Bab	t -∅	t	s	h	h	
Pap	t, s -∅	t, s	s, ∅	h	h	y
Hoa	t, s -∅	t, s	s	h	h	
Sir	s	s		∅	∅	

Table 2: Formosan reflexes of Proto-Austronesian (*cont'd*)

PAN Sq	*m / _#	*ŋ	*n	*ŋ' (*ñ)	*ʔ (*N)	*l	*r
Ata Sq	m -m	ŋ	n	l	l	ɣ, z/_i	
Sk	m -m	ŋ	n	l	l	ɣ	
Mx	m -m	ŋ	n	l	l	∅, w	
Mt	m -m	ŋ	n	l	l	ɣ	
Pl	m -ŋ	ŋ	n	l	l	r	
Sed Tn	m -ŋ	ŋ	n	l	l	r	
Td	m -ŋ	ŋ	n	l	l	r	
In	m -ŋ	ŋ	n	l	l	r	
Tsou	m	ŋ	n, a/a_#	n-h-h	h,k/	r	r
Kan	m	ŋ	n, a/a_#	ŋ	n	l, ∅	r
Sar	m	ŋ	n, a/a_#	ʔ	ʔ	l	r
Ruk Bu	m	ŋ	n	l	l	L	
Mg	m	ŋ	n	l	l	r	r
Mn	m	ŋ	n	l	l	L	r
Bun Tk	m	ŋ	n	n	n	∅	
Is	m	ŋ	n	n	n	∅	
Pai Bu	m	ŋ	n	í	í	L	r
Puy Pn	m	ŋ	n	l	l	L	r
Kl	m	ŋ	n	l	l	L	r
Thao	m	n	n	ǒ	ǒ	r	
Sai Ta	m	ŋ	n	l	l	L	
Paz	m	ŋ	n	l	l	r	
Kav	m	ŋ	n	n	n	r, R/	
Ami Sa	m	ŋ	n	l	ǒ	L	
Fr	m	ŋ	n	l	l	L	
Tao	m	n	n		l	r	r
Bab	m	n	n		l	r	
Pap	m	n	n		l	r, d	
Hoa	m	n	n		l	r, l	r
Sir	m	ŋ	n		l	r, l	

Table 2: Formosan reflexes of Proto-Austronesian (*cont'd*)

PAN	*w	*y	*-uy	*-aw	*-ay, *-əy
Ata Sq	w	y	-uy	-aw	-ay
Sk	w	y	-uy	-aw	-ay
Mx	w	y	-uy	-aw	-ay
Mt	w	y	-uy	-aw	-ay
Pl	w	y	-uy	-aw	-iy
Sed Tn	w	y	-uy	-o	-e
Td	w	y	-uy	-aw	-ay
In	w	y	-uy	-aw	-ay
Tsou	v, Ø	-z-	-uzu	-o, -ou	-e, -oi
Kan	Ø	-l-	-ulu	-a, -au	-ai
Sar	Ø	-t-	-uɬu	-u, -u [?] u	-i, -i [?] i
Ruk Bu	v, Ø/a_i	-d-	-uy	-aw	-ay
Mg	v	-r-	-ivi	-oo	-ee
Mn	v, Ø	-L-	-ui	-ao	-ai
Bun Tk	v	ð	-uð	-av	-að
Is	v	ð	-uð	-av	-að
Pai Bu	v, Ø	-y-	-uy	-aw	-ay
Puy Pn	w	y	-uy	-aw	-ay
Kl	w	y	-uy	-aw	-ay
Thao	w	y	-uy	-aw	-ay
Sai Ta	w	y	-oy	-aw	-ay
Paz	Ø, w	y	-uy	-aw	-ay
Kav	w	y, l	-uy	-aw	-ay
Ami Sa	v	y	-uy	-aw	-ay
Fr	v, w	y	-uy	-aw	-ay
Tao	Ø-				-a
Bab	Ø-	-Ø	-u	-o	-a
Pap		-Ø	-u	-o, -ou	-a
Hoa	Ø-	-Ø	-u	-au	-a
Sir	w	y	-oy		-a

Table 2: Formosan reflexes of Proto-Austronesian (cont'd)

PAN	*a	*i	*u	*ə /_(C)#
Ata Sq	a	i	u	∅, u
Sk	a	i	u	∅, u
Mx	a	i	u	∅, u
Mt	a	i	u	∅, u
Pl	a	i	u	∅, u
Sed Tn	a	i	u	e, u
Td	a	i	u	ə, u
In	a	i	u	ə, u
Tsou	o, a, e, ∅	i, e, ə, ∅, z	u, o, ə, ∅ v	ə, ∅
Kan	a	i	u	ə
Sar	a	i	u	ə
Ruk Bu	a	i	u	ə
Mg	a	i	u	ə, e
Mn	a	i	u	ə
Bun Tk	a	i	u	u, a
Is	a	i	u	a
Pai Bu	a	i	u	ə
Puy Pn	a	i	u	ə
Kl	a	i	u	ə
Thao	a	i	u	i, u
Sai Ta	{ ^a æ/*q, *t', *θ	i	{ ^o œ/*q, *t', *θ	{ ^ə i/*q, *t', *θ
Paz	a	i	u	ə
Kav	{ ⁱ / a/*q	i	u	ə
Ami Sa	a	i	u	ə, a/_*q
Fr	a	i	u	ə, a/_*q
Tao	a	i	u	u, e
Bab	a	i	u	u, ∅
Pap	a	i	u, o	u, e, ∅
Hoa	a	i	u	u, i
Sir	a	i	o, ou	

4.1 The following are the examples of PAN *ʃ (Dahl's (1981a) *S₂) > h in the new Northern group, but s or Ø in the other Formosan languages:

- PAN *kəʃiw > Ata kahu-niq/kahuy, Sed qhu-niʔ, Sai kähöy, Paz kahuy, Bab hau, Pap hé:, Hoa hai, Tso evi, Kan kaalu, Sar kiuʔu, Pai kasiw, Tha ka:viʔ, Sir caiou *tree, wood*, Bun Is busul-kawiʔ *bow* (literally *wood gun*), Ami kasuy *wood*.
- PAN *ku(ʃ)kuʃ > Sed kukuh, Sai ka-k-l-oköh, Bun kuskus, Pai k-al-uskus-an, Tha ku:kuʔ, Ami Sa k-an-us, Ami Fr k-in-uʔus, Kav q-n-uqus *claw, fingernail*.
- PAN *ʃuni > Paz mu-huniʔ, Tso m-uni, Kan um-a-uni, Sar ma-ni-ani, Kav s-m-uniʔ, Ami ma-huniʔ *to chirp*, Tha ʃ-um-a-ʃuni *bird name*.
- PAN *qəʃuŋ > Ata qhuŋ, Sed Tr qihuŋ, Tso uŋo, Kan uuŋu, Sar uʔuŋa, Ruk Mg ŋu-ŋu, Ruk Mn ʔuŋu, Bun quuŋ, Tha quun *mushroom*.
- PAN *Cuʃuy > Ata l-um-uhug (< A- l-/c-)⁸, Sed l-m-ihug (< A- l-/c-), Sai s-om-höL, Kan c-um-acuuru, Bun ma-tusul, Pai c-um-usu, Kav t-m-usuR *to thread a needle, to string beads*.
- PFN *ʃuyəta > Ata hula-qiy, Sed hudaʔ (-d- irregular), Sai hä-höläʔ, Paz ha-həlaʔ, Hoa o-hut-ta, Tso ruxo, Kan əréna, Sar uruta, Ruk Mn ʔola, Pai sula, Puy urla, Tha ʔuṭṭa, Ami suLṭaʔ, Kav suRnaʔ *snow, ice*

As Dahl (1981a:35) pointed out, only Bun has two different reflexes for PAN *ʃ: -Ø- in *kəʃiw and *qəʃuŋ but s in *ku(ʃ)kuʃ and *Cuʃuy. Reflexes in all other languages are quite regular.

4.2 The following are examples of PAN *h (*H₁) > h in the new Northern group and in Bunun Tk and Ami, and Ø in other Formosan languages.

(a) Final Position

Note that Babuza, Papora and Hoanya generally lose final *h; Papora retains final *h only in the form PAN *baʃah > baláh below:

- PAN *baʃah > Ata bagah, Sed bagah, Pazeh bahah (< A- medial h/x), Pap baláh, Kan vara, Sar varaʔa, Ami Fr faLah *charcoal*, Sai baläh, Bab bagga, Sir vaga *burning charcoal*.
- PAN *baqəyuh > Sediq bgu-ra-h, Bab bao, Pap barú, Sir vaho, Tso farv-a, Kan vaʔuru-a, Sar varuʔu, Ruk baav-ánə, Bun baqlu, Pai vaqu-an, Tha faqtu, Ami vaqluh-ay *new*.
- PAN *buŋuh > Tso fŋuu, Kan nə-vuŋu, Sar vuŋuʔu, Bun buŋuʔ, Ami Fr fuŋuh, Sir vongo *head*.
- PAN *qumah > Ata qum-quma-a-h/quma-qumah, Kan ʔúuma, Sar umu-uma, Ruk uma-uma, Bun qumaʔ, Pai quma, Puy ʔuma, Sai ʔöm-ʔömah, Paz uma-umah, Ami umah, Bab ema, Hoa uma, Sir ouma *farm, field*, Tso mo-mo *to work in the field*.
- PAN *Cuṭuh > Ata c-um-uluh, Tso c-m-uhu, Kan c-um-á-cunu, Sar c-um-acuṭu, Bun ma-tunu, Pai culu, Sai s-om-olöh, Ami mi-tuluh *to roast right on the fire*.

PAN *d'atih > Sed dalih, Sai ?äl-?älih-an, Paz ?alih, Kan ará-cani, Sar ma-saŋi, Ruk Mg me-d-dali, Ruk To ma'a-da-diali, Ruk Mn ma'a-ðii-ðali *near*.

PAN *tutuh > Tso m-uutu *to strike or hit (in general)*, Kan m-aká-tutu *to hit with the fist*, Sai totöh *to strike with a stick*, Paz kalu-batu mu-tutuh *to strike with a stone*.

PAN *nunuh > Sed nunuh, Paz nunuh, Tso nun'u, Sar nuunu?-a *breasts*.

(b) Initial and Medial Positions

PAN *bahi > Ruk a-ba-bai, Pai va-vay-an, Puy ba-bay-an, Ami Fr va-vahi?-an *woman*.

PAN *lahud > Tso mua-rovcu *to blow downhill*, moh-rovcu *to flow downstream*, Kan ?ama-laúcu *downhill*, m-a-a-láucu *to blow down the hill*, Sar tala-la-laucu *to look down*, Ruk Mg róDu, Ruk Tn aúDu *downhill*, Ruk Mn lauðu *downwards*, Pai lauz *seaward*, Tha mana-raws *downhill*, Sai Lähör *downhill*, haw-Lähör *downstream*, Paz rahat *west*.

PAN *buhut > Ata bhut, Sed b-ri-huc, Kan vuutu, Ruk buu-buutu, Bun puhut (< A- p/b), Pai vuŋ, Puy but, Sai ka-bhöŋ, Paz buhut, Ami vuhut *squirrel*.

PAN *hulə[tŋ] > Tso rəsə *upper garment*, Bun huluc (< A- ə/u), Kav quŋus, Ami hələt (A- ə/u, irregular -t) *type of clothing*.

(Thao form hu:lus is probably a loanword from Bunun Takbanuað hulus.)

PAN *ha(n)taq > Bun mə-hacaq, Pai t-əm-ataq *to whet*.

4.3 The following are examples of PAN *ŋ > l in the new Northern group:

PAN *qaŋud > Ata mə-qaluit (?), Sed qluli?, Tso ŋ-ohcu, Kan m-a-?acúnu (n/ŋ metathesis and assimilation to c), Sar m-u-aŋusu, Ruk Bđ mu-aluDu, Ruk Mg mu-luDu, Ruk Mn mu-luðu, Bun muŋ-qanu?, Pai sə-qalud, Puy mu-lahud (< M), Ami mə-qalul (A- l/r), Sai lö-?älur *to flow, to be adrift*.

4.4 Many Formosan languages have the lateral fricative /l/[t]⁹ as reflex of PAN *t (Dyen's *N) as well as PAN *ŋ. These languages are Atayal, Sediq, Rukai, Paiwan, Puyuma, Saisiat, Pazeh, Taokas, Babuza, Papora and Hoanya. Formosan examples for PAN *t are as below:

PAN *taŋuy > Ata l-um-aŋuy Tso ru-huŋzu, Kan mə-ka-naŋulu, Sar maka-taŋulu, Ruk wa-laŋuy, Pai l-əm-aŋuy, Puy mə-la-laŋuy, Sai l-om-aŋoy, Paz mu-laŋuy, Kav m-naŋuy, Ami mi-laŋuy, Sir l-m-aŋuy *to swim*, Sed l-m-aŋuy *to bathe in a stream*.

PAN *tibu > Ata libu?, Sed libu?, Kan niívu, Sar tivu?u, Ruk Bđ libu, Ruk Mg libuu, Ruk Mn livu, Pai íivu, Tha ði:fu, Sai libu?, Ami ðibu? *den, nest*, Paz libu-patakan *fence to keep out pigs, etc*.

- PAN *tuwan > Kan ?i-núanə *female deer*, Sar ta-i-tuanə *female pygmy deer*, Bun qa-nvan, Tha q-nu:wan (< A- n/ð) *deer, carabao*, Paz nuan (< A- n/l) *carabao, cow*, Hoa loang *deer*, Sir louang *elk, horse, ox*.
- PAN *təfún > Ata luhin/luhun, Sed duhun (d- irregular), Tso suhu (< M s/h), Sar tuuŋu, Ruk luuŋu, Bun nucuŋ, Puy Lp lusun, Sai löhöŋ, Paz luzun, Kav ?i-nsun *mortar*.
- PAN *Daŋum > Tso chumu, Kan canumu, Sar saŋumu, Bun danum, Pai zaŋum, Puy zanum (< A- n/l), Tha sa:ðum, Sai ralom, Paz dalum, Ami nanum, Kav zanum, Bab dalom, Pap dom, lom, róm, Hoa zazum, radum, Sir salom *water*.
- PAN *qaŋup > Ata q-um-alu-a-p/q-um-alup, Sed m-aduk (irregular -d-), Kan ?-um-a-?anupu, Sar ?-um-a-?aŋupu, Ruk w-áŋupu, Bun qanup, Pai q-əm-alup, Puy Kp ?-əm-alup, Sai ?-öm-alop, Ami mi-qaðup, Sir m-alup *to hunt*, Tso hup-a *hunting territory*.
- PAN *taŋam > Ata t-um-alam, Sed t-m-alaŋ, Tso oo-thomə, Kan paku-tanam-ən, Sar mak-taŋam, Bun tanam, Puy t-əm-alam, Tha t-m-a:ðam, Sai ŋan-talam, Paz mu-talam, Ami mi-tanam *to taste*.
- PAN *bulaŋ > Ata Mx bua-tiŋ, Ata Sk byal-iŋ, Tso frohə, Kan vuánə, Sar vulaŋə, Bun buan, Puy bulan, Tha fu:rað, Ami buLað, Kav vulaŋ, Pap budal, Hoa bulan, Sir vourel *moon*.
- PAN *quzaŋ > Ata qual-ax,¹⁰ Sed qu-yux, Tso m-əchə, Kan ?ucánə, Sar usaŋə, Ruk údalə, Bun qudan, Pai qudaŋ, Puy ?udal, Tha qu:sað, Sai ?ä-?öral, Paz udal, Ami quðað, Kav ?uzan, Tao m-otal, Bab hutas, hutat, Pap m-odal, Hoa m-udzas, Sir oudal *rain, to rain*.

4.5 Reflexes for PAN *C are either c or s in the Northern group of Formosan languages. The same is also found in the Tsouic group, in Rukai and Paiwan. However, it is t in Bunun, Kavalan and Ami, ʔ in Puyuma, θ in Thao. Examples are:

- PAN *Caliŋə > Ata caŋia? (< M), Sar caliŋə, Ruk caŋiŋə, Bun taiŋə, Pai caŋiŋə, Puy TaŋiLa (< M), Tha ʔari:na (A. θ/ʔ), Paz saŋiŋə? (< M), Ami taliŋə?, Tao salina, sareina, Bab harina, charrina, Pap sarina, Hoa saŋina, sarina, sangila, Sir tangira *ear*, Kan caŋiŋə *ear ornament*.
- PAN *maCa > Tso mcoo, Ruk maca, Bun mata?, Pai maca, Puy maŋa, Tha ma:θa, Sai masa?, Kav mata?, Ami mata?, Tao masa, Bab macha, Pap masa, Hoa masa, Sir matta *eye*.
- PAN *kuCuŋ > Ata ku-hiŋ/kucu?, Sed qu-hiŋ (< A- q/k), Tso ?cuu, Kan kúucu, Sar kucu?u, Ruk kucu, Bun kutu?, Pai kucu, Tha ku:θu, Sai koso?, Paz kusu?, Kav qutu?, Ami kutu?, Tao usu, Bab ocho, Pap uθu, Hoa usu *head louse*.
- PAN *huyaC > Ata Sq ?ug-iq, Sed ?urac, Tso vroçə, Kan urácə, Sar ?uracə, Ruk Bd, Mg uváčə, Ruk Mn ?účə, Bun ?ulat. Pai St ?uwac, Puy ?uraŋ, Tha ?u:ʔaθ, Paz huhas (< A- h/x), Ami ?uLat, Pap yolas, yoras *blood vessel, vein*, Sai ka-uLas *sinew*.
- PAN *kuliC > Puy kuliŋ, cf. Yami kulit *bark of tree*.
- PAN *laŋiC > Sar laŋic-a, Puy laŋiŋ *sky*.

5. PHONOLOGICAL EVIDENCE FOR THE INTERNAL RELATIONSHIPS OF THE NORTHERN FORMOSAN LANGUAGES

5.1 The clustering of Taokas, Babuza, Papora and Hoanya

As Tsuchida (1982:9) pointed out, the four sinicised Formosan languages Taokas, Babuza, Papora and Hoanya are more closely related to each other than to any other Formosan languages. These languages share the innovations exclusively of other Formosan languages: (1) PAN *k > Ø (Tsuchida 1982:9), (2) PAN *-y > -Ø (Tsuchida 1982:9), (3) PAN *ŋ > n (Thao is the only other Formosan language which shares this innovation), and (4) PAN *t' > t in initial and medial position (Tsuchida 1982:10), but -Ø in the final position. However, there are exceptions to these rules. There are dialect variations in these languages. Moreover, the language data based on the memory of the few old people may not be very accurate.

(1) Examples for PAN *k > Ø:

- PAN *kuCuḥ > Tao usu, Bab ocho, Pap uθu, Hoa usu *head louse*.
 PAN *kayəŋ > Tao y-axan, Bab aggan, Pap aqan *crab*.
 PAN *kaən > Tao úman, Bab man, Pap man, Hoa man *to eat*.
 PAN *kiṭa > Bab m-ita, Hoa k-am-ita *to see*.
 PAN *kajiw > Bab hau, Pap hé:, Hoa hai *tree*.
 PAN *aku > Tao y-au I.
 PAN *bukəs > Pap bus, bud, buθ, Hoa bud *hair*.
 PAN *pirak > Tao pira, Pap parái *money*, Hoa pira *silver coin*.
 cf. Ata, Sed pila?, Paz parái? *money*.
 PAN *aṭak > Pap da, lah, lala, Hoa ala, ālak *child*
 PAN *sikan > Hoa sikan *fish*.
 PSF *uka > Tao ua *not have*.

(2) Examples for PAN *-y > -Ø:

- PAN *babuy > Bab, Pap, Hoa babu *pig*.
 PAN *maCəy > Bab macha, maha, Hoa maθa *to die*.
 PAN *paCəy > Bab pacha, Pap mada *to kill*.
 PAN *pagəy > Bab adda, Pap pada, Hoa padza, paza *rice plant*.
 PAN *quay > Bab choa *rattan*.
 PAN *bəyəy > Bab pea (?), Pap bula *to give*.

(3) Examples for PAN *ŋ > n:

- PAN *Caliqah > Tao sarina, Bab harina, Pap sarina, Hoa sarina, saninna *ear*.
 PAN *Caŋit' > Tao s-am-ani, Bab s-um-ani, Pap s-am-ani, Hoa s-um-ni, s-m-anē *to weep*.
 PAN *ṭuaŋ > Bab loan, Pap loan, luaŋ, Hoa loan, loang *cow, carabao*.
 PAN *kayəŋ > Tao y-axan, Bab aggan, Pap aqan *crab*.

(4) Examples for PAN *t' > t in initial and medial positions, Ø in final:

- PAN *əṭa > Tao t̄ā-nu, Bab n̄a-ta, Pap ta-nu, Hoa m-eta, itta, a-ta *one*.
 PAN *atú > Bab atu, Pap h-atu, Hoa atu *dog*
 PAN *t'utú > Hoa tutu *breasts (of woman)*.
 PAN *Caŋit' > *to weep* (see (3) above).
 PAN *mamit' > Tao mme, Pap mame, Hoa mami *sweet*.

Although PAN *t' > t is also found in Paiwan and Thao, it is lost in the word-final position only in these four sinicised Formosan languages.

5.2 The clustering of Saisiat and Pazeh

Saisiat and Pazeh exclusively share the following innovation: PAN *g' > z. For example,

- PAN *baqaq > Sai bazä? *to listen*, Paz ma-baza? *to know*.
 PAN *Sagək > Sai s-om-azək, Paz mu-sazək *to smell*.
 PAN *pagəy > Sai pazay *cooked rice, rice plant*, Paz pazay *glutinous rice*.
 PAN *qa(m)pəgú > Sai pä'zo? (< M), Paz 'apuzu? (< A- u/ə) *gall-bladder*.
 PAN *pigəh > Sai piza? *how many*.
 PAN *mugin > Paz muzin *nose*.

However, while Pazeh merges PAN *g' with PAN *t' and *θ (first reconstructed by Tsuchida 1976) to z, Saisiat retains the distinction between PAN *g' > z and PAN *t' and *θ > h. For example,

- PAN *butuy > Sai böhöL *bow*, Paz buzux *arrow*.
 PAN *tətun > Sai löhöŋ, Paz luzun *mortar*.
 PAN *θuθu > Sai höhö? *breasts of woman*.
 PAN *θəpθəp > Paz mu-zəzəp *to suck*.

5.3 The clustering of the Atayalic group

The Atayalic group comprises two main languages, Atayal and Sediq. Atayal, in turn, consists of two major dialect groups, Squliq and C'uli?(Ts'ole). Squliq dialects are all fairly uniform, whereas C'uli? dialects can be quite divergent from each other and some of them preserve some interesting and archaic features. These dialects include Ba'ala?, Bay'anux, Mabatun, Mnawyan, Maspazi?, Matabalay, Mayrinax, Palıawan, Pianan, Skikun, Sakuxan and Talawan. Sediq has relatively uniform dialects, including Toŋan, Toda and Inago (Taroko). There are both phonological and lexical differences between these dialect groups. I shall not go into the details of these differences in this paper; interested readers are referred to Li (1980a, 1981).

The most unique feature in the entire Atayalic group is that there are male and female forms of speech in every dialect. See Li (1980b, 1982c) for a detailed account.

The following phonological changes are shared by all dialects in the Atayalic group:

1. PAN *ə > u/_(C)#

PAN *ə has become u in all Atayalic dialects in the final syllable except in the diphthong *-əy, and this change is unique to the group. For example, verbs with the suffix *-ən, as in *bəlig'-ən, indicating the Referential Focus has become -un, as in Sq bir-un, Sk bes-un, Mx binas-un, Pl binaf-un, Td briw-un, In brig-un *to buy*. Other examples are:

- PAN *lusəq > Sed rusuq *tears*.
 PAN *qayəm > Ms 'agum, Mx qagum, Sed 'aruŋ *anteater*.
 PAN *tələn > Sq hayun, Mx hawun, Sed harun *pine tree*.
 PFN *tuqəs > Sq luqus, Ms lu'us, Mx luqus *brains, marrow*.

cf. PAN *pagəy > Ata pagay, Sed payay *rice plant*.

2. PAN *t', *θ, *ʃ, *h > h

The Atayalic group has merged PAN *t', *θ, *ʃ and *h as h, and the same change is found only in Saisiat. However, the Atayalic has a different reflex for PAN *h (Dahl's (1981a) *H₂), which has become Atayalic ʔ, Saisiat Ø and -h.

(1) PAN *t' > h

PAN *batuq > Sq, Sk mahuq, Ms mahuʔ, Mx ma-bahuq, Pl ma-bahuʔ,
Sed mahuʔ *to wash clothes*.

PAN *batuy > Ms bah-in-uw, Sk bh-en-ux, Mx buh-in-ug, Pl bahu-niʔ,
Tn bhe-naq, Td bhə-niq *bow*.

(2) PAN *θ > h

PAN *əθi > Ata hiʔ, Sed hiiʔ *meat, flesh*.

(3) PAN *ʃ > h

PAN *kafuy > Sq, Sk, Mn qho-niq, Mx kahu-niq/kahuy, Mb kahu-niq,
Ms kahaw-niʔ, Mt, Pl kahu-niʔ, Sed qhu-niʔ *tree, wood*.

PAN *ʃakuC > Sq, Bl h-m-akut, Sk h-m-akuc, Mx, Mb, Mt h-um-akut, Ms,
Sx h-am-akut, Pl h-um-akuc, Sed h-m-akul (-l is a sporadic
change, cf. PAN *cumah > Sk, Mx lum-iq *body louse*) *to carry*.

(4) PAN *h > h

PAN *buhut > Sq bhut, Ms bahut, Sk bhuc, Mx bhut, Pl buhuc, Sed
b-ri-huc *squirrel*.

PAN *qumah > Sq, Sk, Mn qma-ya-h, Ms ʔama-ya-h, Mx quma-qumah/
qum-quma-a-h, Mb, Mt, Sx ma-ma-ya-h, Pl muma-ra-h, Bx ma-ya-h
cultivated field.

3. PAN *d, *D, *d' > Ata r, Sed d/Initial and Medial
Ata t, c, ʔ, Sed c/Final

The Atayalic group has merged PAN *d, *D and d' as Atayal r and Sediq d in initial and medial position, and as Atayal (most dialects) ʔ, Ata Mt t, Ata Pl and Sediq c in final position:

PAN *dayaq > Sed daraʔ *blood*.

PAN *daqis > Ata Sq rqi-ya-s, Mx raqi-na-s, Sed dqə-ra-s *face*.

PAN *Daya > Ata Sq k-ryaʔ, Sk rayaʔ, Mx maka-ryaʔ, Sed dayaʔ *inland, up*.

PAN *dalan > Ata Sk ryan-iq, Mx ran-iq/raan *road*

PAN *dáyum > Ata raum/ragum *needle*.

PAN *dátih > Sed dalih *near*.

PAN *paŋdaʔ > Ata paŋraŋ *pineapple*.

PAN *quDas > Ata quriʔ/quras, Sed Td qudas *gray hair*

PAN *qañud > Ata Sq m-qluiʔ, Mt ma-qaluit, Sed qluliʔ *flow*.

PAN *paʔiD > Ata paliʔ, Sed palic *wing*.

PAN *taʔiuD > Sq, Sk tliuʔ *mulberry*.

PAN *qəluD > Sed ʔəruc *pillar*.

In addition, there is the sporadic change PAN *k > Atayalic q when followed by q or h perhaps by assimilation; see Li (1980:377) and Li (1981:247-248).

Examples are:

PAN *kəən > Ata qən-iq *Eat!*

PAN *kuCuh > Sed qu-hiq *head louse*

PAN *kafuy > Ata Sq qhu-niq, Sed qhu-niʔ *tree, wood*

6. LEXICAL EVIDENCE

All known Austronesian cognates are excluded from the following lists. Such cognates are at best only weak evidence for subgrouping even if they are retained only in these languages.

6.1 Cognates that are exclusively shared by Atayalic, Saisiat, Pazeh, Taokas, Babuza, Papora and Hoanya

1. Ata tutuh, Paz tutu?, Tao tutoh, Pap duduh *tobacco pipe*.
2. Ata Sq musa?, Ata Mx ma-usa?, Sed In musa?, Paz mu-ku-sa?, Tao muha, m-a-usa, Bab musa, Hoa usa, muha *to go*.
3. Ata, Bab papak *ear*.
4. Ata, Sed In basag, Sai basal, Tao básau *millet*.
5. Ata Mx balayan, Sai balayan, Paz bulayan, Tao burayan, Pap buleyan *pan for cooking*.
6. Ata, Paz patus, Pap patos *gun*.
7. Sed l-lebu?, l-ləbu?, Pap ma-lbu *low*.
8. Ata Sq miru? (biru?), Paz mu-bizu? *write*, Bab bido *ink*, Pap ba-bidu, Hoa ma-bidzu *tattoo*.
9. Ata suwai?, Sed swai?, Paz swazi?, Hoa suazi *younger sibling*.

6.2 Cognates that are exclusively shared by Atayalic, Saisiat and Pazeh

The Atayalic languages share the following cognates with Saisiat and Pazeh, but not with any other non-sinicised Formosan language. Some of these may be due to borrowing.

1. Ata Mx, Paz syatu? *clothes, upper garment*.
2. Ata Mx qaim, Ata Ms ?azim-ux, Sai ?äLim, Paz ?arim *peach*.
3. Ata Ms yapuwar, Sai lapwar, Paz lapwat *guava*.
4. Ata ?ali?, Paz ?ali? *bamboo shoots*.¹¹
5. Ata ?abag, Paz rabax *leaf*.
6. Ata tu-nux *head*, Sai tonu? *brains*, Paz tunu? *brains, marrow*.
7. Ata Ms baun, Sed In ba?un, Paz baun *pumpkin*.
8. Ata, Paz taba? *gourd*.
9. Ata Ms, Paz Kh tami? *sponge gourd*.
10. Ata, Sai tabuil *cucumber*.
11. Ata tatupun, Paz tatupun *type of mushroom*.
12. Ata, Sai kamti? *dye yan*.
13. Ata Sq zimala?, Ata Ms, Pl yamala?, Paz rumala? *plant sp. (Diplazium esculentum (Retz.) Swartz)*.
14. Ata basiyaw, Sed Td bsiyaw, Paz basyaw *plant sp. (Alpinia speciosa Schum)*.
15. Sed bruwa?, Sai bilwa? *thunder*.
16. Ata Mx balayan (< A a/u), Sai balayan (< A a/u), Paz bulayan *cooking pan*.
17. Ata waqa-nux, Sai wä?ä? *deer*.
18. Ata Mx bauwak, Ata Ms bayuwak, Ata Pl barok, Paz baruzak *pig*.
19. Ata Sk, Mx, Paz batu? *egg*.
20. Ata, Sed kui?, Paz kui? *worm*.
21. Ata kabah-niq, Sed qbh-eni?, Sai kab-kabäh-äl *bird*.

22. Ata parnah, Paz paranah *boat*.
23. Ata qa-qibug, Sai ʔä-ʔiboL, Paz Kh ʔayxuʔ *paddle*.
24. Ata Mx ruas, Sai z-in-ulas *paper*.
25. Ata Mx ginuʔ, Paz xinuʔ *winnowing basket*.
26. Ata Sk, Sai, Paz takil *women's basket which is carried on one's back*.
27. Ata Mx, Paz salaman bowl.
28. Ata Mx, Sai, Paz yasam *ax*.
29. a. Ata ca-capin/ca-capuh, Sai sa-sapöh, Paz sa-sapuh *broom*.
b. Ata c-um-apin/c-um-apuh, Sai s-om-apöh, Paz mu-sapuh *to sweep*.
30. Ata malahan, Sed qmlahan, Sai malahan *to take care of*.
31. Ata Mx ma-tauaw, Ata Mt ma-tayuaw, Sai mataLoaw *to work*.
32. Ata Ms rarumaʔ, Sai raromäh, Paz rarumaʔ *type of bamboo*.

6.3 Cognates exclusively shared by Saisiat, Pazeh and four north-western sinicised Formosan languages (Taokas, Babuza, Papura and Hoanya)

1. Paz tulalaʔ, Bab tullala, Pap, Hoa tulala *flower*.
2. Sai bonaz, Paz, Tao bunat, Bab bonnad *sand*. cf. Ata bunaqiy, which is ambiguous with Ami bunak *id*.
3. Sai baLalaʔ, Tao baxada, Pap varada *river*.
4. Paz baugul, Tao baugun *frog*.
5. Paz turak, Pap turák *penis*.
6. Sai Lasəb, Paz xasəp, Tao xasap, Bab nahup, nachap, Hoa hasip *five*.
7. Paz ʔisit, Tao ta-isid, Bab tsixit (?), Pap (me-)tsii, Hoa myata-isi, Luilang isit, Fav¹² zchiett *nine*.
8. Paz Kh habak, Bab abak *boat*.
9. Sai tiʔäʔ, Tao tíxaʔ, Bab tsia, Pap lidaq *needle*.
10. Paz parəpar, Hoa pa-lat-pa *paper*.
11. Paz laluzuk, Bab lallokogh, Pap laloqol (metathesis of the last two consonants?) *comb*.
12. a. Sai tanoLan, Tao tanran *pail*.
b. Paz gunugun, Pap gungun, Hoa gun-gun *pail (wooden tub)*.
13. Sai wa-watos, Tao watos *gun*. cf. No.17 below.
14. Paz ka-kazip, Hoa kazib *chopsticks*.
15. Paz ʔaluʔ, Bab alo, Pap aro *Come!*
16. Paz, Hoa mahatan *to laugh*.
17. Paz pa-patus, Tao patus, Pap patuθ *to fire a gun*. cf. No. 13 *gun* above.
18. Paz maliak, Bab malleak *to dislike*.
19. Paz mulasiʔ, Tao malasi *rice plant*.
20. Paz sumay, Tao suma, Bab sma, Hoa smai *cooked rice*.
21. Sai, Paz tawtaw, Tao tau-tau *peanut*.
22. Paz yamaduʔ, Tao yamadu *sugar*.
23. Paz ka-kumus, Tao kumus, Pap komū *hat*.
24. Paz lasuʔ, Bab, Hoa lasu *husked rice*.
25. Paz sibabun, Bab tebabon, Pap chibabun, Hoa bababun *duck*.
26. Sai ʔälaw, Paz ʔalaw, Bab alau *fish*.

6.4 Cognates exclusively shared by Saisiat and Pazeh

1. Sai ʔə-tiLim *mote in the eye*, Paz tirim-ən *to have mote in the eye*.
2. Sai Laməs, Paz xaməs *root*.
3. Sai ra-rLaŋ, Paz daxaŋ *sweat*.
4. Sai laliw, Paz ma-laliw *earthquake*.
5. Sai pöhäk, Paz puhak *bubble*.
6. Sai, Paz lakay *a wild fruit like mango*.
7. Sai kaway, Paz ka-kaway *a bag or basket which is carried on one's back*.
8. Sai bilis, Paz mu-bilis *to hold*.
9. Sai h-om-azab, Paz mu-hazap *to stab*. cf. Ata Mx h-um-ab *to stab*.
10. Sai Lilafjan, Paz xilasan *male pheasant*.
11. Sai lihjay (< D l/h), Paz zihikay *maggot*.

6.5 Cognates exclusively shared by Taokas, Babuza, Papora and Hoanya

1. Tao budum, Bab boesum, Pap burom, Hoa budzum *sky*. cf. Sir vullum *id.*
2. Bab, Hoa pisi, Pap pisī *monkey*.
3. Tao hibin, Bab ibien *snake*; Pap, Hoa lisu *id.*
4. Tao, Bab mura, Hoa mula *face*. cf. Sir moula *id.*
5. Tao tilax, Bab tatsira, Pap tatsiá, tatina *tongue*. cf. Sir dadila *id.*
6. Tao yanut, Bab, Pap, Hoa nut *nose*.
7. Tao yudah, Bab oda *intestines*.
8. Tao ttaho, Bab ttao *hair*; Pap, Hoa bud *id.*
9. Tao, Pap mabasak *blind*.
10. Bab, Pap gyobe *mouth*.
11. Tao ridok, Pap idok *breasts (of women)*.
12. Tao, Hoa tiyat *belly*.
13. Tao rrapyi-masa, Pap rralpī *eyebrow*.
14. Tao, Hoa baha *meat*.
15. Tao, Bab, Pap, Hoa tapaha *trousers*.
16. Bab silok, Hoa siluk, Sir silock *sword*.
17. Bab kaiyu, Pap kaiʔyú:, Hoa kayu *bowl*.
18. Bab aismala, s-um-ala, Pap mala, mlat, Hoa malat *to sleep*.
19. Bab mappē, Pap mappe, Hoa mapi *bitter*.
20. Bab matsis, Pap matsit, Hoa machit *salty*.
21. Tao, Pap mari, Hoa mali *sour*.
22. Tao ye-tadas, Bab ma-tadach, Pap ma-dadas, Hoa ma-das(t) *hot (of weather)*.
23. Pap vuhun, Hoa buhun *neck*.
24. Tao matah, Bab mataha, Pap matata, Hoa mataha *blue*.
25. Tao kakan, Bab ma-kakan *red*.
26. Tao marom, Pap mahom *yellow*.
27. Bab baas, Pap baθ *year*.
28. Tao hai-sanat, Bab he-sanas, Pap sanat, Hoa saⁿgat *star*.
29. Tao taanu, Pap tanu *one*.
30. Bab nuxan *cooked rice*, Hoa m-uhan *to eat cooked rice*.
31. Tao tāpu, Bab ta-apu *father*.
32. Bab, Hoa nai *mother*; Bab, Pap kaya *id.*
33. Tao, Bab laugu *grandfather*.

34. Tao ririm, Bab jijim *small child*.
35. Bab tamimi, Hoa tamimi, tamami *sweet potato*, cf. Sir tamamy *id*.
36. Bab hapi, Pap hapi? *areca nut*.
37. Pap, Hoa lun *jacket, coat*.
38. Tao yev, yayeb, Bab eeb *tree, fire-wood*.
39. Tao málók, Bab malok *dog*.
40. Bab bottos, Hoa butūs *cat*.
41. Tao tipan, Bab tsipan *west*.
42. Tao madaish, Bab madich *sick*.
43. Tao papat, Bab ma-pappa *short*.
44. Pap, Hoa majen *beautiful*.
45. Tao tahos, Bab ma-ta-och *heavy*; Pap manil, Hoa manen *id*.
46. Bab ch-um-ap, i-chap, Pap s-um-ap *to close (the door)*.

As stated in Tsuchida (1982:9), "Among these [sinicised] languages, there seems to be a clustering of Taokas and Babuza on the one hand, and of Papora and Hoanya on the other." The lexical evidence given above seems to indicate that that is the case: the clustering of Taokas and Babuza as in No. 3, 7, 8, 25, 31, 33, 34, 38, 39, 41, 42, 43 and 45, and the clustering of Papora and Hoanya as in No. 3, 8, 44 and 45. It is clear that evidence for the former clustering is stronger, whereas evidence for the latter is weaker.

Siraya, another sinicised language, appears to share some cognates with the four languages under discussion as in No. 1, 4, 5, 16 and 35. However, there is little phonological or lexical evidence for grouping Siraya with the Northern group.

NOTES

1. This paper was written with the support of the National Science Council Grant NSC70-0301-H001-08, Republic of China. I wish to thank Andrew Pawley, Isidore Dyen and Pang-hsin Ting for comments on this paper.
2. The following is a list of abbreviations of the Atayalic languages and dialects:

Ata	Atayal	Mt	Matabalay	Sq	Squliq
B1	Ba?ala?	Mx	Mayrinax	Sx	Sakuxan
Bx	Bay?anux	PA	Proto-Atayalic	Td	Toda?
In	Inago	Pl	Palɔawan	Tl	Talawan
Lm	Lmuan	Pn	Pianan	Tn	Toŋan
Mb	Mabatu?an	Ps	Piasan	Tr	Truwan
Mn	Mnawyan	Sed	Sediq		
Ms	Maspazi?	Sk	Skikun		

3. The male and female forms of speech are separated by the slash "/" and are in that order.
4. Some reconstructions for Formosan languages are based on Tsuchida (1976). PFN stands for Proto-Formosan and PSF for Proto-Southern-Formosan.

5. The hypothesis of the new subgroup was first suggested by Dahl (1981b:55 and personal communication, 22 October 1982).
6. There is an exception: Saisiat distinguishes between *ni* *genitive marker for personal names (including 'father' and 'mother')* and *some pronouns* and *no genitive marker for non-personal proper noun*; see Li (1978:600).
7. Abbreviations of the Formosan languages and dialects except the Atayalic group (see Section 2, note 2) are as given below. Language names are listed in Column 1, and dialect names in Column 2 in Table 1:

Tso	Dh	Tsou	Duhtu	Sai	Ta	Saisiat	Taai
Ruk	Bu	Rukai	Budai	Paz		Pazeh	
		Mg	Rukai	Kav		Kavalan	
		Mn	Rukai	Ami	Sa	Ami	Sakizaya
Bun	Tk	Bunun	Takituduh		Fr	Ami	Faraɔaw
		Is	Bunun	Tao		Taokas	
Pai	Bu	Paiwan	Butanglu	Bab		Babuza	
Puy	Pn	Puyuma	Pinan	Pap		Papora	
		Lp	Puyuma	Hoa		Hoanya	
		Kl	Puyuma	Sir		Siraya	
Tha		Thao					

Citations in this study are taken from the first dialect of each language unless stated otherwise. Atayal is generally based on the Mayrinax dialect. I suspect that Taokas dialect No.10 listed in Tsuchida (1982) is actually the Mayrinax dialect of Atayal and so I have excluded it from Taokas.

8. Sporadic changes are indicated in the parenthesis, e.g., < A- \bar{p} /b means that b has changed to p by assimilation; M indicates metathesis; D indicates dissimilation.
9. The voiced lateral fricative [ɬ] is represented by the symbol /l/, while the voiceless lateral fricative [ɬ̥] is represented by /ɬ/ in this study.
10. If the Atayal form qual-ax is cognate, then the medial consonant *-Z- > -r- is inexplicably lost.
11. Bun Tk (in the village of Kukuað) ?ali *bamboo shoots* is probably a loan from Atayal and/or Pazeh.
12. It is clear from lexical evidence that Favorlang (Happart 1650) is a dialect of Babuza, and Tsuchida (1982:17) treats it as such in his monograph.

BIBLIOGRAPHY

BLUST, Robert

- 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics, University of Hawaii* 9/2:1-15.

CHEN, Teresa M.

- 1985 *Verbal constructions and verbal classification in Nataoran-Amis*. (PhD dissertation, University of Hawaii at Manoa, 1982.) *PL*, C-85.

DAHL, Otto Christian

- 1973, *Proto-Austronesian*. Scandinavian Institute of Asian Studies,
1976 Monograph Series No. 15. Lund: Studentlitteratur.
- 1981a *Early phonetic and phonemic changes in Austronesian*. Oslo: Institute for Comparative Research in Human Culture.
- 1981b Austronesian numerals. *NUSA* 10:46-58. Jakarta.

DEMPWOLFF, Otto

- 1934- *Vergleichende Lautlehre des austronesischen Wortschatzes*. *ZES*
1938 Beihefte 15, 17, 19.

DYEN, Isidore

- 1965a *A lexicostatistical classification of the Austronesian languages*. Indiana University Publications in Anthropology and Linguistics Memoir 19, supplement to *IJAL*.
- 1965b Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285-305.
- 1971a The Austronesian languages and Proto-Austronesian. In Sebeok, ed. 1971:5-54.
- 1971b The Austronesian languages of Formosa. In Sebeok, ed. 1971:168-199.

EGEROD, Søren

- 1965a Verb inflection in Atayal. *Lingua* 15:251-282.
- 1965b An English-Atayal word list. *Acta Orientalia* 19:203-220.
- 1966a Word order and word classes in Atayal. *Language* 42/2:346-369.
- 1966b A statement on Atayal phonology. *Artibus Asiae Supplementum XXIII* (Felicitation volume for the seventy-fifth birthday of Prof. G.H. Luce), vol.1:120-130.
- 1980 *Atayal-English dictionary*. Scandinavian Institute of Asian Studies, Monograph Series 35. London: Curzon Press.

FERRELL, Raleigh

- 1969 *Taiwan Aboriginal groups: problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica, Monograph 17. Taipei.

FERRELL, Raleigh

- 1979 Construction markers and subgrouping of Formosan languages. In Nguyễn Đăng Liêm, ed. *South-East Asian linguistic studies*, vol.3: 199-211. PL, C-45.
- 1979 Phonological subgrouping of Formosan languages. In Naylor, ed. 1979:241-254.

HAPPART, Gilbertus

- 1650 Woord-boek der Favorlangsche Taal, ed. by W.R. Van Hoëvell. *Verhandelingen van het Bataviaasch Genootschap* 18:31-381.

HARVEY, Mark

- 1980 Subgroups in Austronesian. M.A. thesis, The Australian National University, Canberra.

JENG, Heng-hsiung

- 1977 *Topic and focus in Bunun*. Institute of History and Philology, Academia Sinica Special Publication 72. Taipei.

LI, Paul Jen-kuei

- 1973 *Rukai structure*. Institute of History and Philology, Academia Sinica Special Publication 64. Taipei.
- 1977 Morphophonemic alternations in Formosan languages. *Bulletin of the Institute of History & Philology, Academia Sinica* (hereafter BIHP) 48/3:375-413.
- 1978 The case-marking systems of the four less known Formosan languages. In S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 569-615. PL, C-61.
- 1980a The phonological rules of Atayal dialects. *BIHP* 51/2:349-405.
- 1980b Men's and women's speech in Mayrinax. *Papers in honor of Professor Lin Yü-k'eng on her seventieth birthday*, 9-17. Taipei: Wen Shin Publishing Co.
- 1981 Reconstruction of proto-Atayalic phonology. *BIHP* 52/2:235-301.
- 1982a Atayalic final voiced stops. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 171-185. PL, C-75.
- 1982b Linguistic variations of different age groups in the Atayalic Dialects. *Tsing Hua Journal new series* 14:167-191.
- 1982 Male and female forms of speech in the Atayalic group. *BIHP* 53/2: 265-304.

MARSH, Mikell Alan

- 1979 The Pazeh-Saisiat-Favorlang subgroup of Formosa. In Naylor, ed. 1979:255-264.

NAYLOR, Paz B., ed.

- 1979 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*. Ann Arbor: The University of Michigan.

OGAWA, Naoyoshi

1931 *Japanese-Atayal dictionary*. Taipei.

OGAWA, Naoyoshi and Erin ASAI

1935 *Myths and traditions in the Formosan native tribes*. Taipei: Taihoku Imperial University.

SEBEOK, Thomas A., ed.

1971 *Current trends in linguistics, vol.8, Linguistics in Oceania*, 5-54. The Hague: Mouton.

STAROSTA, Stanley

1974 Causative verbs in Formosan languages. *Oceanic Linguistics* 13:279-369.

1985 Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou. In this volume, pp. 281-312.

TSUCHIDA, Shigeru

1976 *Reconstruction of Proto-Tsouic phonology*. Tokyo: Study of Languages & Cultures of Asia & Africa, Monograph Series No.5.

1982 *A comparative vocabulary of Austronesian languages of sinicized ethnic groups in Taiwan Part I: West Taiwan*. Memoirs of the faculty of Letters, University of Tokyo No.7.

TUNG, Tung-ho

1964 *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology, Academia Sinica Special Publication 48. Taipei.

WURM, Stephen A. and B. WILSON

1975 *English finderlist of reconstructions in Austronesian languages (post-Brandstetter)*. PL, C-33.

YAMADA, Yukihiro and Ying-chu LIAO

1974 A phonology of Tayal. *Research Reports of the Kochi University* 23/6:109-117.

YANG, Hsiu-fang

1976 The phonological structure of the Paran dialect of Sediq. *BIHP* 47/4:611-706.

VERBAL INFLECTION VERSUS DEVERBAL NOMINALISATION IN PAN: THE EVIDENCE FROM TSOU

Stanley Starosta

1. INTRODUCTION

In a paper presented at the Third International Conference on Austronesian Linguistics in Bali in 1981, Starosta, Pawley, and Reid (hereafter SPR) proposed a scenario for the evolution of Philippine-type clause structure in which, at a stage ancestral to Philippine and certain other languages, PAN equative sentences with nominalised NP predicates were reinterpreted as verbal sentences. However, the question has been raised by Shelly Harrison (personal communication) as to how far this process had already gone at the stage SPR reconstruct as Proto-Austronesian, and how much of the preexisting verbal syntax, as opposed to SPR's competing nominalised constructions, we are able to reconstruct for PAN.

Part of the answer to this question is suggested by the syntax of the Tsouic languages of Taiwan, and in particular in the structure of Tsou itself. Tsouic languages form a very high order subgroup of the Austronesian language family, possibly a primary subgroup, yet Tsou does not have the very strong nominal orientation which SPR reconstructed for PAN. Of the five derivational affixes they considered to be crucial in the evolution of Philippine-type focus constructions, *mu-/-um-, *-en, *ni-/-in-, *-ana, and *iSi-, Tsou has clear reflexes of only two, *mu-/-um- and *-ana, with only *mu-/-um- involved in verbal constructions.¹ In place of the ubiquitous nominalised attribute and nominalised predicate constructions of the Atayalic and Paiwanic languages of Formosa, Tsou complex NP constructions are composed of a head relator noun and a sentential attribute which is unmistakably verbal in its syntactic properties.

Based on a comparison of Tsou and the other Formosan languages, it turns out to be possible to reconstruct the stages that led to the modern Tsou focus system, but only if we either

1. revise the SPR view of the nature of PAN clause structure,
2. show that the PAN noun-derived focus marking system could plausibly have been lost in Tsou while being retained in the other Tsouic languages, or
3. revise current assumptions about the way that Tsou subgroups with the other Tsouic languages, Saaroa and Kananavu, and with Atayalic to the north and Rukai to the south.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 281-312. *Pacific Linguistics*, C-88, 1985.

© Stanley Starosta

281

2. THE TROUBLE WITH TSOU

Starosta, Pawley, and Reid posited a PAN system in which nominalised equational constructions had to some undetermined extent been reinterpreted as verbal constructions in main clauses by analogy with a preexisting verbal focus system involving focus affixes *-a and *-i². Since Tsou verbal constructions reflect only one of these presumed nominal affixes, *mu-/-um-, and otherwise employ the supposedly earlier *-a and *-i for Object Focus and Locative Focus, SPR are presented with a problem. Whether Tsouic is a primary PAN subgroup, as proposed by Harvey (Figure 1) and Reid (Figure 2), or a primary branch of Proto-Southern Formosan (i.e. non-Atayalic), as proposed by Tsuchida (Figure 3), it should reflect the denominal verbal focus affixes in verbal constructions, since all its sisters and cousins do. (Actually, the situation in Rukai is quite similar to that in Tsou, but I will not consider it further in this paper.)

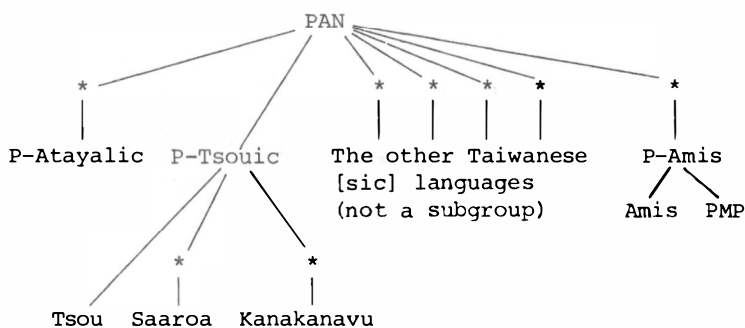


Figure 1 (based on Harvey 1979:98, 104)

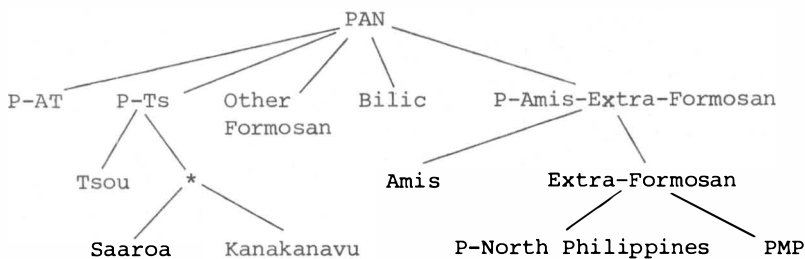


Figure 2 (based on Reid 1981:15)

There are three basic options open to us in accounting for the Tsou situation. Assume that either

1. the use of originally nominal affixes in verbal clauses was independently innovated in all languages but Tsou (cf. the asterisks in Figures 1, 2 and 3), or
2. Tsou originally had the same set of affixes but lost them, or
3. the subclassification trees are wrong, and Tsou itself is a primary subgroup, with all the other languages allocated to different subgroups or forming a single subgroup (Figure 4):

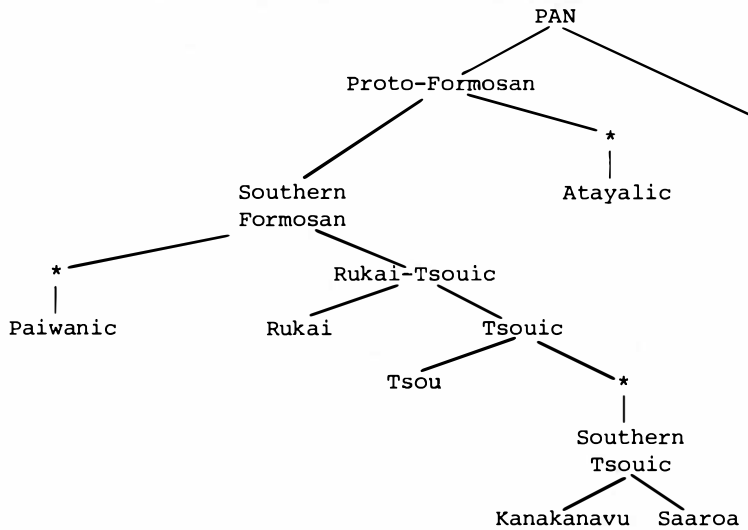


Figure 3 (based on Tsuchida 1976:13, 15)

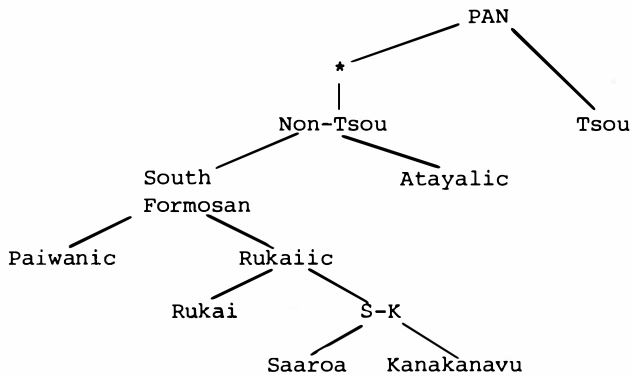


Figure 4

Option 1 is not a very attractive one. While the reinterpretation of nominals as verbs is itself not at all an implausible change, as SPR tried to show³, it would still be surprising to have exactly the same process happening independently in exactly the same way in so many different branches, at the points indicated by the asterisks in the family tree diagrams shown as figures 1, 2, and 3.

Option 3 is counterindicated by the comparative studies which have been done so far, since none of these put Atayalic together with Paiwanic as opposed to Tsou.

In terms of economy and plausibility, then, it would be desirable to work toward the second alternative. That is, it would be nice if SPR could assume that the ancestors of Tsou did have the denominal verbal affixes in question

and then show how Tsou could have lost them (cf. Wolff 1973:74). That is what I propose to attempt in the main body of this paper.

3. AUX CLUTCHING

According to SPR, the innovating verbal focus affixes in PAN itself were only present in main clauses, with the original *-a and *-i focus affixes preserved in subordinate clauses. This is of course the situation preserved to varying degrees in Austronesian languages such as Seediq (Asai 1953:28) and Samar Leyte (Wolff 1973:87), and is in fact the situation SPR reconstructed for the ancestor of many of the Oceanic languages. Moreover, SPR claimed that auxiliary verbs were syntactically the highest verbs of their respective clauses in PAN, as they are in the modern languages (cf. Ross 1969, Starosta 1977), so that when an auxiliary cooccurred with another verb, the second verb would have been syntactically a subordinate verb, and thus have occurred with the earlier subordinate clause focus affix set.

An example from English may help to illustrate this point. From the point of view of Ross's 'Auxiliaries as main verbs' analysis, the highest verb in an English sentence such as John must leave for Parramatta soon is not leave but must, so that the bracketing would be:

(_S John must (_S leave for Parramatta soon))

where must is the highest tensed finite verb of the sentence, and leave is the infinitival head of the embedded complement of must. Assuming this kind of analysis and the SPR account of PAN clause structure, then, all non-auxiliary verbs cooccurring with auxiliary verbs would have been syntactically subordinate, and thus would have appeared with the *-a and *-i focus affixes.

Given these two features of PAN syntax, we could in principle explain the development of the Tsou system in terms of a single innovation if we could somehow motivate the requirement that every sentence contain an auxiliary verb. I will refer to such a development as 'Aux clutching'. Aspect-marking auxiliary verbs must have been common in PAN, since they are also very frequent in Atayalic as well as Tsou, and it is in fact a striking feature of Tsou syntax that almost every verbal sentence in connected discourse contains an initial aspect-marking auxiliary verb (Tung's 'beginners'; Tung 1964:88-89). This auxiliary verb is frequently followed by a clitic pronoun (Tung's 'post-beginners' of the /si/ group; Tung 1964:89), and these clitic pronouns occur *only* immediately after auxiliary verbs.⁴ These clitics are unusual in terms of Formosan languages in the requirement that they coreference actors rather than subjects (cf. Tung 1964: 100, 107-109).⁵ In the following section, I will attempt to show why and how 'Aux clutching' transpired in Tsou as a new mechanism for marking aspect, mood, and pronominal actors.

3.1 Actor-referencing clitic pronouns

It is possible in each of the Tsouic languages to distinguish a set of independent pronouns as well as a Nominative and a non-Nominative clitic set (Tsuchida 1976:38, 68, 98; Mei 1982:209), though in Tsou itself the latter distinction in the clitic system is clear only in the third person singular (Tsuchida 1976:97). As usual in Formosan languages and elsewhere, the case

form of the clitic is crucial in determining coreference: Nominative clitics coreference the grammatical subject in the next clause down, and non-nominative clitics coreference non-Nominative actants in the embedded clause. This situation is illustrated in figures 5, 7, and 9. Illustrative Tsou examples are given after each of the schematic tree diagrams. Bracketed numbers refer to sentence numbers in my field notes:

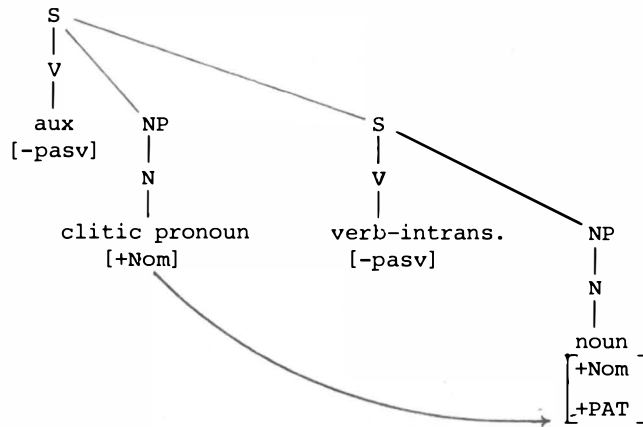


Figure 5 Intransitive clause: schematic

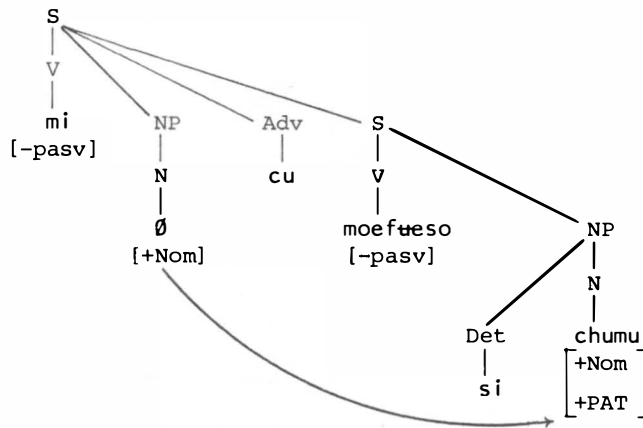


Figure 6 Intransitive clause

(Note that, by the lexibase Patient Centrality hypothesis, every intransitive verb has a Patient subject; cf. Bruce 1983.) More examples:

- (1) mi cu moefueso si chumu [C25]
 1 2 3 4 5
The water has already boiled.
 4 5 1 2 3

- (2) mo na?no ~~umnt~~ si oko [C39]
 1 2 3 4 5
The child is very good.
 4 5 1 2 3
- (3) mo eo ne ~~fueŋt~~ o mameoi [C36]
 1 2 3 4 5 6
The old man is in the mountains.
 5 6- -6 1- -2 2-3 4
- (4) te ko n?a uh ne oeonatmopst [C161]
 1 2 3 4 5 6
(You will now) go to school!
 2 1 3 4 5 6
- (5) mo ~~macht~~ mai tan?e [C23]
 1 2 3 4
It rained today.
 1-2 3-4
- (6) mi ?o sɔ?no [C19]
 1 2 3
I got angry.
 2 1-3 -3
- (7) mi ?o oefɔ?ɔ [C27]
 1 2 3
I fell down.
 2 1-3 -3

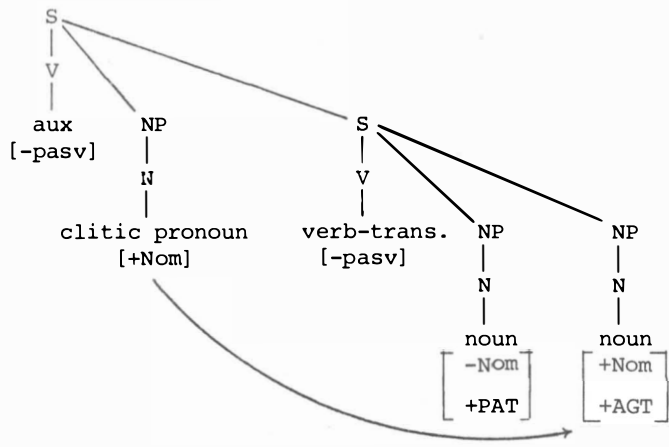


Figure 7 Transitive active clause: schematic

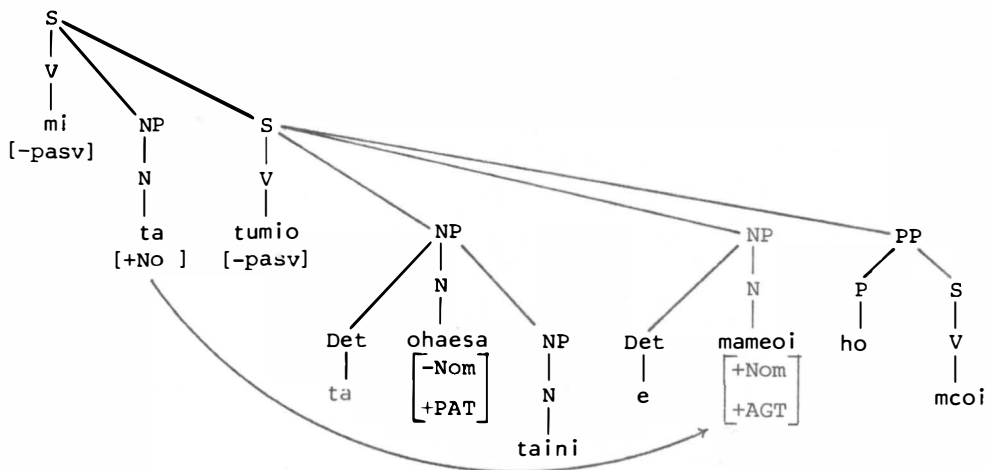


Figure 8 Transitive active clause

- (8) mi ta tumio ta ohaesa taini e mameoi ho mcoi [C22.3]
 1 2 3 4 5 6 7 8 9 10
The old man died for his younger brother;
The old man benefited this younger brother to die.
 7 8- -8 3 6 5- -5 9 10
- (9) mio eobako to fkoi o mameoi [C45]
 1 2 3 4 5 6
The old man hit/killed a snake.
 5 6- -6 2 3 4
- (10) mo pei?i ta chumu si mamespiŋi [C86]
 1 2 3 4 5 6
The woman boiled (cooked) water.
 5 6 2 - 1 4
- (11) mi ?o bon to simeo [C13]
 1 2 3 4 5
I ate the fat meat.
 2 1-3 4 5- -5
- (12) mo ?u cu mofi to peisu to mamespiŋi [C155]
 1 2 3 4 5 6 7 8
I gave the woman the money.
 2 1-3 7 8 5 6

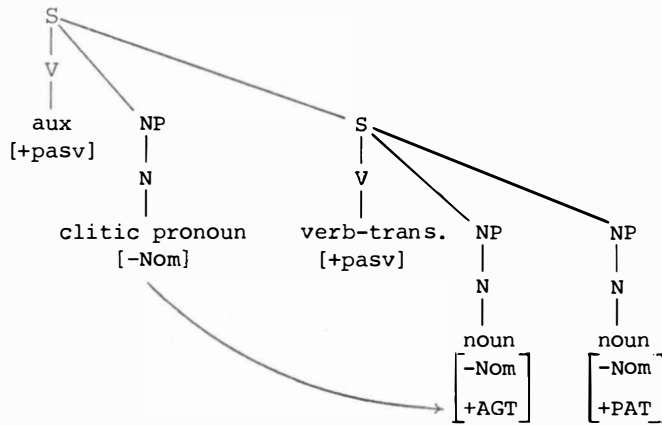


Figure 9 Transitive passive clause: schematic

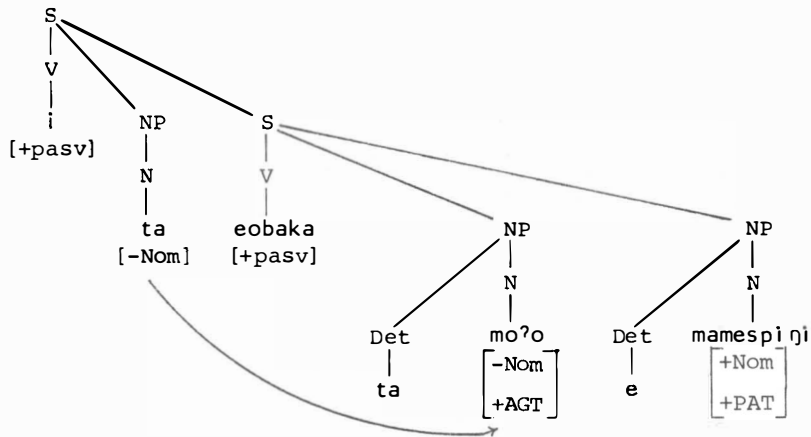


Figure 10 Transitive passive clause

- (13) i ta eobaka ta mo'o e mamespiŋi [C153]
 1 2 3 4 5 6 7
The woman was hit by Moe.
 6 7 1-3 -3 4 2-5
- (14) i ta tufkuneni ta poeoeo ta oko e mameoi [C8.2]
 1 2 3 4 5 6 7 8 9
The child washed the old man's pants for him;
The old man was washed pants for by the child.
 8 9- -9 1 3- 5 -3 6 2-7
- (15) i'o mameoi i si fii to mo goen o oko ne hucma [C1]
 1 2 3 4 5 6 7 8 9 10 11 12
Yesterday the old man gave the child five dollars;
The old man, the child was given five dollars by him yesterday.
 1 2- -2 9 10 3 5 7-8- -8 4 11-12

- (16) i'e oko i ta eobakneni ta ab'u e mameoi [C41]
 1 2 3 4 5 6 7 8 9
This child beat the dog for the old man;
The child, the old man was beaten the dog for by him.
 1 2 8 9- -9 3 5- 6 7 -5 4- -4
- (17) a os 'o cohivi a te uh ne fuen+ si mameoi [C41]
 1 2 3 4 5 6 7 8 9 10 11
I know the old man will go to the mountain.
 3 4 10 11- -11 6 7 8- -8 9
- (18) io i si pei'i ta vcon-si ci naaveu na'no mafe [C59]
 1 2 3 4 5 6 7 8 9 10 11
The food which his wife cooks is very delicious.
 8 9 1 7 6 4 10 11
- (19) i si cu poa-moefueso ta mamespiŋi si chumu [C87]
 1 2 3 4 5 6 7 8 9
The woman made the water boil;
The water was made to boil by the woman.
 8 9 1 4 5 6 7-2
- (20) i si poa-moeai to si'ŋi ta amoo-si o oko-si [C121]
 1 2 3 4 5 6 7 8 9 10 11 12
The father had his children make a broom;
 7 8-2 3 12 11 4 5 6
His children were caused to make a broom by their father.
 12 11 1 3 4 5 6 9 8-2
- (21) i si poa-mofia ta mameoi to peisu o oko-su [C105]
 1 2 3 4 5 6 7 8 9 10 11
The old man told your child to bring the money;
 5 6-2-6 3 11 10 4 7 8
Your child was caused to give the money by the old man.
 11 10 1 3 4 7 8 5 6- 2-6
- (22) i si poa-bon+ ta mameoi to fou o ino-si [C126]
 1 2 3 4 5 6 7 8 9 10 11
The old man had his mother eat meat;
 5 6-2-6 3- 11 10 -4 8
His mother was caused to eat the meat by the old man.
 11 10 1 3 4 7 8 5 6-2-6
- (23) i si poa-an-eni ta mameoi ta maaea si simeo [C129]
 1 2 3 4 5 6 7 8 9 10 11
The Japanese had the old man eat fat meat;
 8 9-2 3 6 7- -7 4 11- -11
The fat meat was caused to be eaten by the old man by the Japanese.
 10 11- -11 1 3 5 4 6 7- -7 8 9
- (24) os 'o s+?nova e mamespiŋi [C80.1]
 1 2 3 4 5
I hate/got mad at this woman;
The woman was gotten angry at by me.
 4 5 1 3- -3- -3 2
- (25) os 'o fii ta peisu e mamespiŋi [C157]
 1 2 3 4 5 6 7
The woman was given the money by me.
 6 7 1 3 4 5 2

- (26) te to nʔa eobaka o fkoi [C162]
 1 2 3 4 5 6
Let's go beat the snake;
The snake will now be beaten by us.
 5 6 1 3 4- -4 2
- (27) os ʔo eobakneni ta fatu si kaapana [C47]
 1 2 3 4 5 6 7
I strike the bamboo against the stone;
The bamboo is struck against the stone by me.
 6 7 1 3- -3 4 5 2
- (28) te ko nʔa poa-fae-neni to amoo-su to mameoi o peisu [C96]
 1 2 3 4 5 6 7 8 9 10 11 12 13
Tell your father to give the money to the old man;
 4 9 8 5 12 13 10 11- -11
The money will now be had given to the old man by you.
 12 13 1 3 6 4 5 10 11- -11 2
 (Literal passive glosses will be omitted for the remaining imperative examples)
- (29) te ko nʔa poa-eobako to avʔu o moʔo [C111]
 1 2 3 4 5 6 7 8 9
(You will now) tell Moe to beat the dog.
 2 1 3 4 9 5 6 7
- (30) te ko nʔa poa-moeai no siʔŋi o oko-su [C119]
 1 2 3 4 5 6 7 8 9 10
(You will now) tell your child to make a broom.
 2 1 3 4 10 9 5 6 7
- (31) te ko nʔa poa-mofi to mameoi to peisu o amoo-su [C97]
 1 2 3 4 5 6 7 8 9 10 11 12
(You will now) tell your father to give the money to the old man.
 2 1 3 4 12 11 5 8 9 6 7- -7
- (32) to ko nʔa poa-mofia to peisu o oko-su [C104]
 1 2 3 4 5 6 7 8 9 10
(You will now) tell your child to bring/give the money.
 2 1 3 4 10 9 5 6 7
- (33) i si poa-faeni-neni ta oko to amoo-si to mameoi o peisu [C99]
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
The father made his child give money to the old man;
The money was had given to the old man by the child.
 13 14 1 3 4 5 11 12- -12 6 2-7
- (34) te ko nʔa poa-eobak-neni to moʔo o avʔu [C115]
 1 2 3 4 5 6 7 8 9 10
(You will now) tell Moe to beat the dog.
 2 1 3 4 8 5 9 10
- (35) te ko nʔa poa-an-eni ta maaea si simeo [C128]
 1 2 3 4 5 6 7 8 9 10
(You will now) give that fat meat to the Japanese to eat.
 2 1 3 4 9 10- -10 7 8 5

- (36) i si poa-eobak-neni ta oko to mo'o o av'u [C118.2]
 1 2 3 4 5 6 7 8 9 10 11
Moe had the child hit the dog;
The dog was caused by Moe to be hit.
 10 11 1 3 8 9-2 5 4
- (37) te ko n'a eomia tan'e ho teai si sapiei [C61.4]
 1 2 3 4 5 6 7 8 9
(You will now) use this to repair the shoes.
 2 1 3 4 5 6 7 8 9
- (38) te ko n'a eomia tan'e ho mooeai (ta) sapiei [C61.4]
 1 2 3 4 5 6 7 8 9
(You will now) use this to make (the) shoes.
 2 1 3 4 5 6 7 8 9

Note that a subclass of Tsou aspect-marking auxiliaries may be marked for passive,⁶ and that auxiliary verbs are subject to the Tsou requirement that embedded verbs agree in passivity with the matrix verb.

The factor that crucially distinguishes Tsou from its sisters in terms of clitic coreference behaviour is that Tsou clitics can only coreference the 'actor' of the lower clause (cf. Tung 1964:100, 107-109), where 'actor' is used in the Role and Reference sense (cf. Foley 1976: Abstract, p.2, and Harvey 1979:39).⁷ That is, Tsou does not allow clauses in which the clitic pronoun coreferences the Patient of a transitive clause, even though the case form is the same:

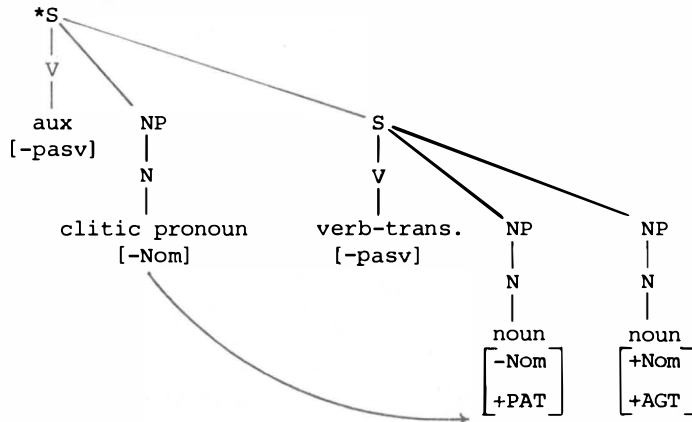


Figure 11 Transitive active clause, Patient clitic

The structure above should be well-formed if clitic coreference operated purely in accordance with case form, since both the clitic and the Patient NP are [-Nom].

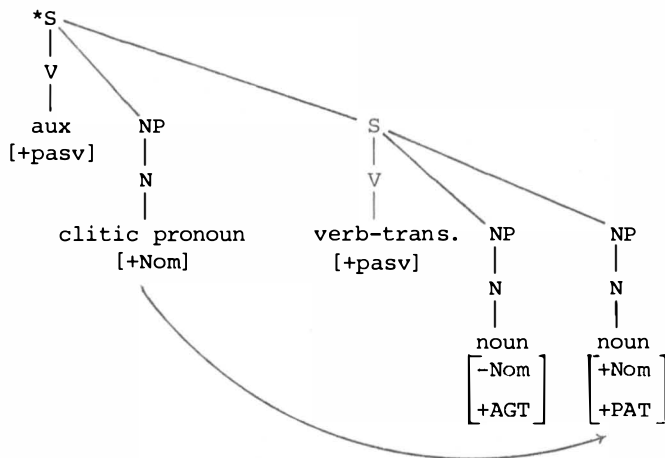


Figure 12 Transitive passive clause, Patient clitic

Again, this structure should be acceptable if clitic coreferencing depended on case form, since the clitic and the Patient share the feature [+Nom]. However, this structure too is impossible in Tsou.

This does not mean, however, that Patients can never be pronominalised. Rather, it simply means that they cannot be pronominalised by means of clitic pronouns. Instead, a separate class of syntactically independent pronouns must be used. This is illustrated in Figures 13 and 15:

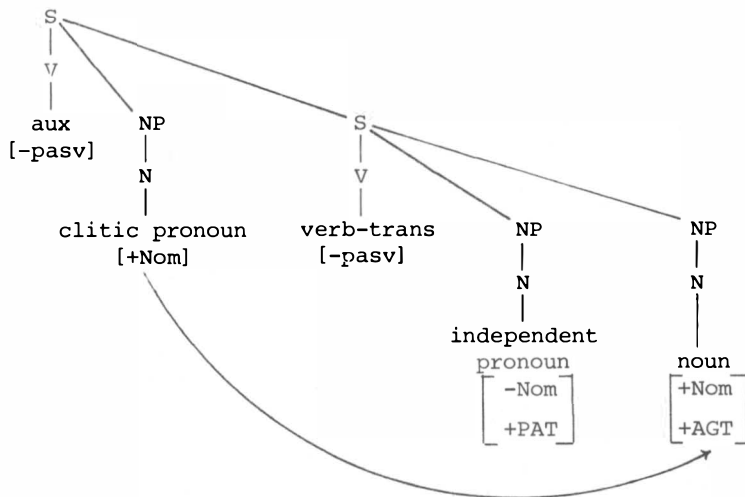


Figure 13 Transitive active clause: schematic

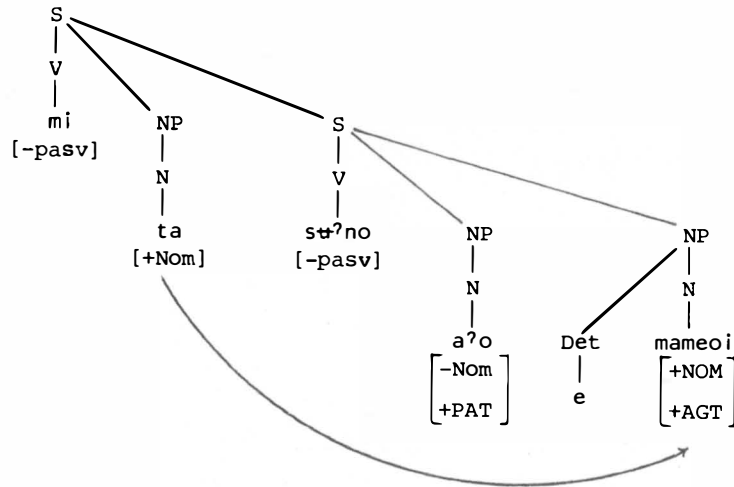


Figure 14 Transitive active clause

- (39) mi ta sɬʔno aʔo e mameoi [C18.11]
 1 2 3 4 5 6
The old man got angry at me.
 5 6-2-6 1-3- -3- -3 4
- (40) mi ta eobako hinʔi [C148.2]
 1 2 3 4
He (visible) hit them (visible).
 2 1-3 4
- (40) *mi si eobako hinʔi [C148.2]
 [-Nom]
- (41) mio eobako suu [C158]
 1 2 3
They (invisible) hit you.
 1-2 3
- (42) mi ʔo eobako suu [C138]
 1 2 3 4
I hit you.
 2 1-3 4
- (43) mi ko eobako aʔo [C136]
 1 2 3 4
You hit me.
 2 1-3 4
- (44) mi ko eobako taini [C140]
 1 2 3 4
You hit him (visible).
 2 1-3 4
- (45) mo hinʔi eobako suu [C158.1]
 1 2 3 4
They (visible) hit you.
 2 1-3 4

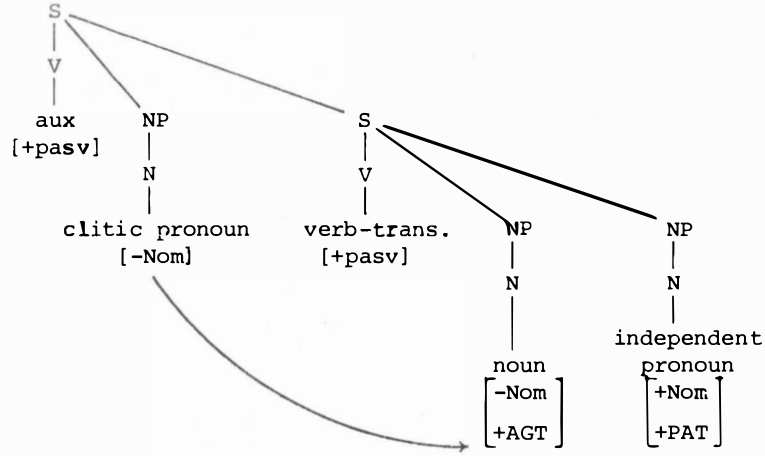


Figure 15 Transitive passive clause: schematic

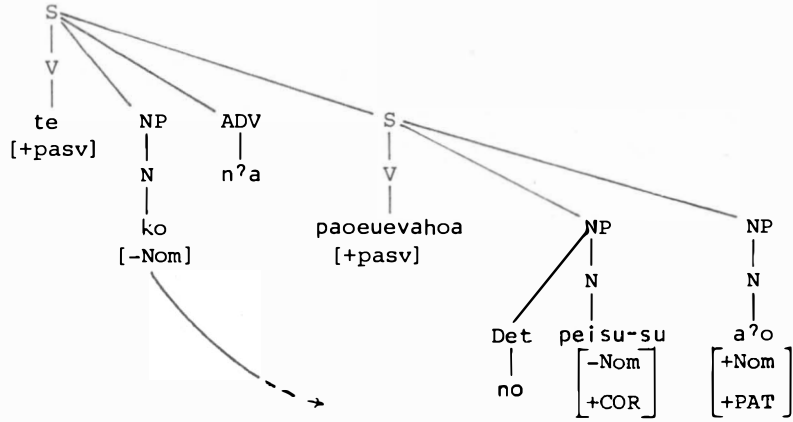


Figure 16 Transitive passive clause

- (46) te ko n?a poaeuevaho(a) no peisu-su a?o [C107]
 1 2 3 4 5 6 7 8 9
Lend your money to me;
I will now be caused to borrow your money by you.
 9 1 3 4 5 8 7 2
- (47) i si eobaka ta?e suu [C141]
 1 2 3 4 5
You were hit by him.
 5 1 3 2-4
- (48) os ?o eobaka taini [C141]
 1 2 3 4
He (close by) was hit by me.
 4 1 3 2

- (49) os ʔo eobaka tonoi [C141]
 1 2 3 4
He (yonder) was hit by me.
 4 1 3 2
- (50) os ko eobaka na aʔo [C137]
 1 2 3 4 5
I was hit by you.
 5 1 3 2
- (51) o ʔu eobaka mu [C147]
 1 2 3 4
You (pl) were hit by me.
 4 1 3 2
- (52) i he eobaka (na) aʔto [C163]
 1 2 3 4 5
We (inclusive) were hit by them. (no information content)
 5 1 3 2
- (53) i mu eobaka aʔo [C151.6]
 1 2 3 4
I was hit by you (plural).
 4 1 3 2
- (54) i si poa-euevaho(a) ta oko to peisu-si aʔo [C109]
 1 2 3 4 5 6 7 8 9 10
*The child had his money lent to me; the child lent his money to me;
 I was caused to borrow his money by the child.*
 10 1 3 4 9 8 5 6-2
- (55) i si pao-euevah-neni aʔo ta oko o peisu-si [C108]
 1 2 3 4 5 6 7 8 9 10 11
*The child lent his money to me;
 His money was caused to be borrowed by me by the child.*
 11 10 1 3 5 4 6 7 8-2

What we have, then, is a system in which clitic pronouns coreference actors, and independent pronouns specialise in representing (transitive and intransitive) Patients. The question naturally arises, then, as to what happens in intransitive sentences with pronoun subjects, since according to lexicase theory, the subject of such sentences would be simultaneously actor and Patient.

We have seen already in Figure 5 that the clitic pronoun can coreference the intransitive subject, which is as it should be, since the intransitive subject is considered an actor. However, if independent pronouns represent Patients, we might rather expect that the subject of an intransitive clause, in its capacity as Patient, should be represented as an independent pronoun. That is, the lexicase Patient Centrality hypothesis suggests that in an intransitive sentence, the single actant could be represented simultaneously by an independent pronoun by virtue of being a grammatical Patient and by a clitic pronoun by virtue of its being an actor, and in fact such sentences are found (Tung 1964:109); e.g.:

- (56) mi ʔo sɬʔno aʔo [C20]
 1 2 3 4
I got angry.
 2,4 1-3 3

This sentence is of course redundant, in the same way that a Spanish sentence with a subject pronoun is redundant: the person and number of the subject is already marked on the (head) verb, which in Tsou is the auxiliary. As in Spanish, though, I think (56) above must also carry an emphatic reading that would not be present if the a^ʔo were absent, as it is in the following example:

(57) mi ʔo s^ʔno [C19]
 1 2 3
I got angry.
 2 1-3 3

These examples can be analysed in a lexicase framework as shown in Figures 17 and 18 respectively:

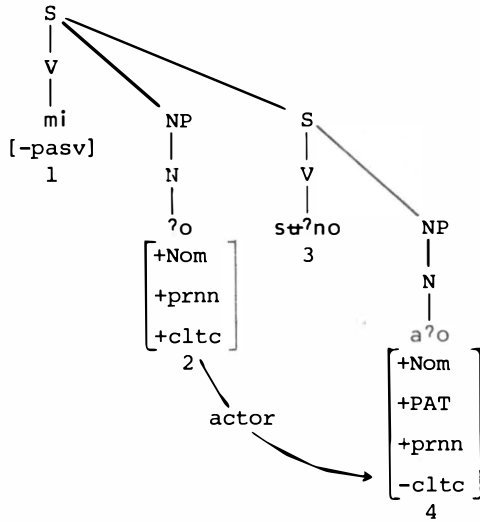


Figure 17 Intransitive clause with coreferential clitic and independent pronoun

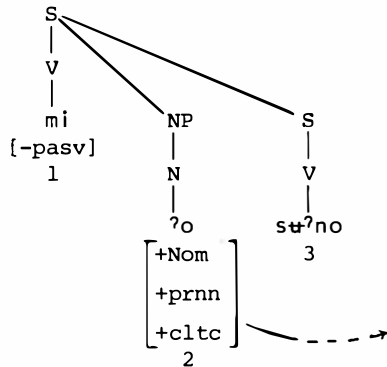


Figure 18 Intransitive clause with coreferential clitic and no independent pronoun

It follows that structures with independent pronoun agents are not grammatical, as illustrated in Figures 19 and 20, identical to Figures 7 and 9 except for the presence of independent Agent pronouns:

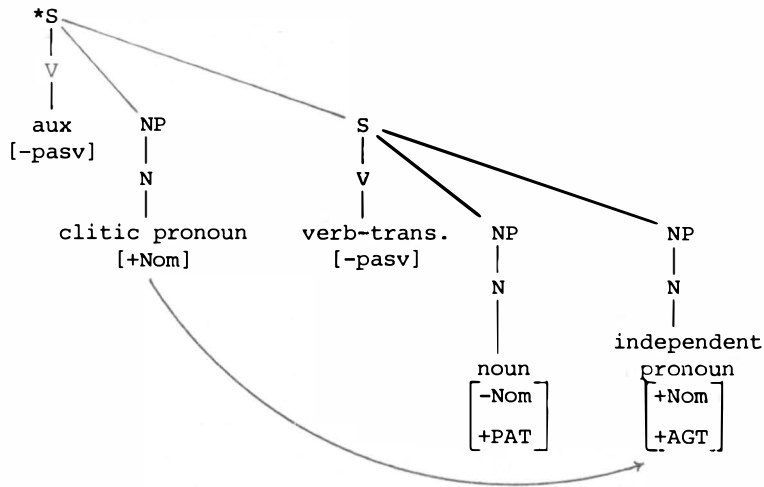


Figure 19 Transitive active clause

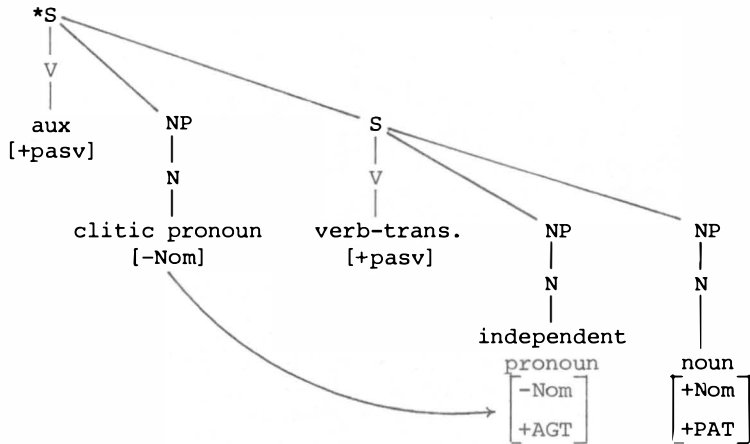


Figure 20 Transitive passive clause

The absence of structures in Tsou like the ones represented in Figures 19 and 20 shows that we can't account for the facts simply as a matter of redundancy avoidance. The structures represented by Figure 17 and Figure 19 are both redundant, but the one represented by Figure 17 is grammatically well-formed and the one represented by Figure 19 is not. This conclusion follows if we assume the correctness of

1. the lexibase hypothesis that every clause has a Patient, and
2. the claim that Tsou independent pronouns are never marked for Agent.

This specialisation of clitics to actors and other pronouns to Patients must be a Tsou innovation,⁹ since Nominative clitic pronouns in Saaroa, Kanakanavu, and Rukai, for example, may coreference lower-clause Patients, of transitive verbs, as shown by examples 58-62:

- (58) taru-cuvuŋ-ani +aku Saaroa (Tsuchida 1976:80)
 1 2 3 4
Come and see me!
 1-2- -1-2 4
- (59) tia-kasu itarə-n Kanakanavu (Mei 1972:213)
 1 2 3
I'll wait for you.
 1 3- -3 2
- (60) vua-(a)u-kani-kia piratiʔiŋi tia kaŋ-a Kanakanavu (Mei 1972:213)
 1 2 3 4 5 6 7
Give me a little of the food.
 1-2 4 5- -5 6-7
- (61) ki-a-baə-ako sa ʔulay Rukai (R9.4)
 1 2 3 4 5 6
I was given to the child.
 4 1-2 3 1 5- -5 6
- (62) ay-ki-θiŋal-su sa umas Rukai (Li 1973:196)
 1 2 3 4 5 6
You will be discovered by a man.
 4 1 2 3 5 6

The -ani in the Saaroa example is a Locative Focus imperative suffix and +aku is a Nominative clitic pronoun, as are the -kasu and -kia in Kanakanavu and the -ako and -su in Rukai. Both Saaroa and Rukai are subgrouped with Tsou as members of the Rukai-Tsouic subgroup, according to Tsuchida, and since both of them behave like the other Formosan languages I have surveyed in this respect, it must be Tsou itself that made the change.

3.2 Clitics, auxiliary verbs, and the loss of finite verbal focus affixes

The effect of this actor-clitic requirement, which was then a purely Tsou innovation and did not affect its sister Rukai-Tsouic languages, is to make an auxiliary obligatory (that is, to make the verb clutch its Aux) whenever the speaker wants to pronominalise an (animate) actor. This includes zero pronominalisation, since it seems that whenever a full Agent or Patient NP actant is missing from a clause in Tsou, an auxiliary must be present as the head verb of the next clause up. With such an auxiliary even the absence of a clitic pronoun is significant, since it unambiguously represents a third person invisible actor in the next clause down, as in (Tung 1964:98): mi cə bəŋu to fʔue (*He*) *baked a sweet potato*, where the auxiliary mi is not followed by any clitic pronoun, and no overt expression corresponds to the implied invisible actor. (bəŋu = *baked*, fʔue = *sweet potato*, cu = *completion*). The nature of discourse is of course such that actors are frequently not represented by full NP's, and as a result, almost every sentence in natural discourse which involves an inanimate actor is preceded by an auxiliary.

The consequences of the appearance of Aux axing in Tsou are quite significant for the thesis of this paper. What we have in effect is a situation in which every non-auxiliary transitive verb which takes an animate actor is embedded under an auxiliary verb, and therefore only takes the dependent set of focus affixes.⁹ Recall that the newly derived finite verbal focus affixes *-en, *ni-/in-, *-ana, and *iSi- initially appeared only in main clauses, so if most transitive clauses have auxiliary verbs, these affixes will no longer appear in most main or subordinate clauses. As the frequency of such affixed verbs decreases, the language increasingly comes to rely on auxiliary verbs to express aspect, and this in turn helps to spread the obligatory auxiliary requirement to intransitive clauses as well. The result is the loss of the main clause set of focus affixes.

Note, however, that *mu-/um- was not included in the list of focus affixes mentioned in the last paragraph. This is because reflexes of *mu-/um- are not lost in Tsou, and in fact are very productive. Thus we must assume that either

1. these affixes had been innovated into the dependent verb paradigm prior to Aux clutching, or
2. they were perhaps never nominal to begin with and did not participate in the main-clause reanalysis proposed in SPR. This latter possibility may in fact turn out to be the correct one, since the evidence for the nominal origins of *mu-/um- has always been somewhat weaker than that for the other focus affixes. This would mean that examples of reflexes of *mu-/um- used as agentive nominalising affixes in modern languages were more recent formations constructed by analogy with the other focus affixes.

4. THE FATE OF DEVERBAL NOMINALISATION IN TSOU

The other problem with Tsou syntax as compared to the system reconstructed for PAN is the prominence and productivity of lexically nominalised constructions in PAN as reconstructed in SPR and the total absence of such constructions in Tsou. PAN and many of its daughters in Formosa and the Philippines make heavy use of the original main-clause focus affixes in nominalised equational constructions in emphatic cleft sentences, content interrogatives, and relative clauses, but Tsou doesn't. In place of the usual equational construction, Tsou has a quite different nominalisation strategy.

Nominalisation among the Paiwanic (non-Tsouic and non-Atayalic) Formosan languages is a process of deriving lexical nouns from verbs by means of the affixes just mentioned. These derived nouns are then used in noun phrases with a range of attributes comparable to the set with which their verbal counterparts can occur. In Tsou, however, one looks in vain for such constructions. What one finds instead in the usual places where nominalisations are expected (relative clauses, cleft equational sentences, and content interrogatives) are constructions composed of one of the case-marking elements, especially na, followed by a modifying clause which is clearly verbal in all syntactic and morphological respects, including the presence of aspect-marking auxiliary verbs. Atayalic, the third major grouping of Formosan languages, exhibits both kinds of constructions, with the deverbal nominalisation strategy stronger in the Sediq subgroup (which is spoken in an area which happens to be adjacent to two Paiwanic languages, Amis and Bunun) and the case-marker-plus-clause strategy stronger in Atayal proper (which has no Paiwanic neighbours except possibly Saisiyat).

Examples of the case-marker-plus-clause strategy of nominalisation for the various functions mentioned above are given below from Tsou:¹⁰

Noun phrases in normal sentential functions

- (63) mo mafe~~est~~ NP [o_S [os ?o eobaka]] (C49)
 1 2 3 4 5 6
The place I hit was very slippery; literally,
What was hit by me is very slippery.
 3 4 6 5- -5 2- -2
- (64) te ko n'a poateaineni NP [o_S [mo kuzo] to ca?h~~+~~] (C124)
 1 2 3 4 5 6 7 8 9
Tell him to repair that broken chair!; literally
You will just repair the is-bad one which is a chair.
 2 1 3 4 5- 6 7 -5 8 9
- (65) NP [i?o_S [te si ana ta oko] ci fou] na?no mafe (C68)
 1 2 3 4 5 6 7 8 9 10
The meat that the child wants to eat is very delicious;
That which will be eaten by the child which is meat, it is very
 1- -1 2 4- -4 5- -5 6-3 7 8 10- 9
delicious.
 -10

Content interrogatives

- (66) NP [sia] NP [na+_S [m-o+ m~~+~~ fr~~+~~?s~~+~~]] Tsuchida 1976:102
 1 2 3 4
Who is the one who covered?
 1 2- -2- -2 3-4
- (67) NP [cuma] NP [na+_S [+ko fr~~+~~?s-a]] Tsuchida 1976:102
 1 2 3 4 5
What is the one which was covered by you?
 1 2- -2- -2 5 4 3
- (68) NP [cuma] NP [na+_S [+ko fr~~+~~?s-eni]] Tsuchida 1976:103
 1 2 3 4 5
What is the one which was covered with by you?
 1 2- -2- -2 5- 4 -5 3
What did you cover with? What did you use for a cover?

Cleft equational constructions

- (69) NP [ina_S [i si seosisi] no teese] Tung 1964:80
 1 2 3 4 5 6
 ma NP [ina_S [la as~~+~~noepoh~~+~~ ho la fei~~+~~na] ci kuhku]
 7 8 9 10 11 12 13 14 15 16
So, what was fastened with the rope was just the creature
 7 1 1-4 4 5 6 8
with whom she had always been together all the night, which
 9 10 11 -11 14 15
was a wildcat.

- (70) NP [io_S [o ho la ahoza aiti] ci oa eatatan'e] si armu Tung 1964:78
 1 2 3 4 5 6 7 8 9 10 11
The Dutch were the first non-natives to be seen.
 10 11 1 5 8 9 6- -6
- (71) ...NP [ie nia ŋohoo] NP [e_S [i to opcoza]] Tung 1964:78
 1 2 3 4 5 6 7
... it is Ngohoo who was killed by us.
 1 3 4 5 7 6- -6
- (72) ...NP [ic'o] NP [na_S [oh to ei'mi]] Tung 1964:81
 1 2 3 4 5
... that is where we live
 1 2 4 3-5
- (73) ...NP [ina_S [ho la rainca honte]] Tung 1964:81
 1 2 3 4 5
 NP [ina_S [moso nana la muro uh to taivaŋ]]
 6 7 8 9 10 11 12 13
... it was one named Honte who came to Taiwan first.
 1 4 5 6 11 13 10

Although it is possible for Tsou equational sentences to occur with initial auxiliary verbs, the cleft equationals normally don't, presumably because the auxiliary verbs inside the nominalised clauses carry all the necessary specifications about aspect.

The constructions exemplified above are problematic from a theoretical point of view: if Tsou case-marking 'particles' are Determiners rather than nouns, what can we make of constructions such as *o os 'o eobaka* *the place I hit* in example 63, which functions syntactically as a subject Noun Phrase but whose two immediate constituents are (i) a supposed Det and (ii) a clearly verbal clause, complete with auxiliary verb and clitic pronoun?

In a transformational framework, of course, one is free to create a dummy head noun and then delete it (Figure 21 below), or to create a new node such as Schachter's NOM (Schachter 1976:206) to serve as the head of the construction (Figure 22), or even to just allow Noun Phrases with no nominal heads at all (Figures 23-25; cf. e.g. Horn 1975:338 and Baker 1978:143ff. for examples of the latter), possibly inserting the determiner by means of a transformation:

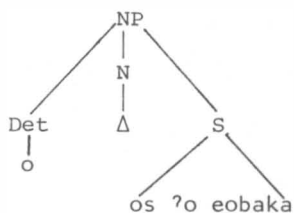


Figure 21

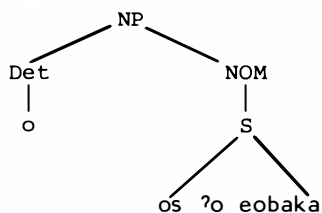


Figure 22

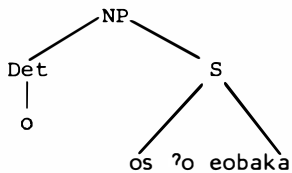


Figure 23

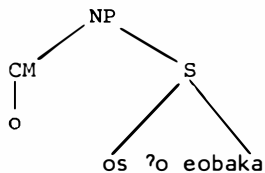


Figure 24

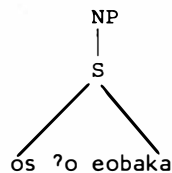


Figure 25

Within the tight constraints proposed by lexicase, though, all of these analyses and many others possible within transformational frameworks and other frameworks of comparable power are ruled out; a Noun phrase must by definition have a lexical noun as its head, and since it is not possible in lexicase to just create one as needed and later delete it, the 'Determiner' is the only plausible candidate. That is, if the 'Determiner' is the lexical head of a noun phrase, then it is by definition a noun.

Accordingly, I would like to propose here that these constructions are themselves relative clauses, and that the initial case-marking elements, in these examples at least, are syntactically nouns which function as the heads of the NP as a whole. That is, they are relator nouns, nouns with minimal semantic content which function as the syntactic head of a construction and carry syntactic or semantic features characterizing the Noun Phrase as a whole (cf. Starosta 1982).¹¹ The following example illustrates this analysis:¹²

- (74) *Tell me where your parents are;* Tung 1964:62
Tell that which it is occupied by them of your parents.
 1 2- -2 4-8 3 6 5--5-7 9 11 1

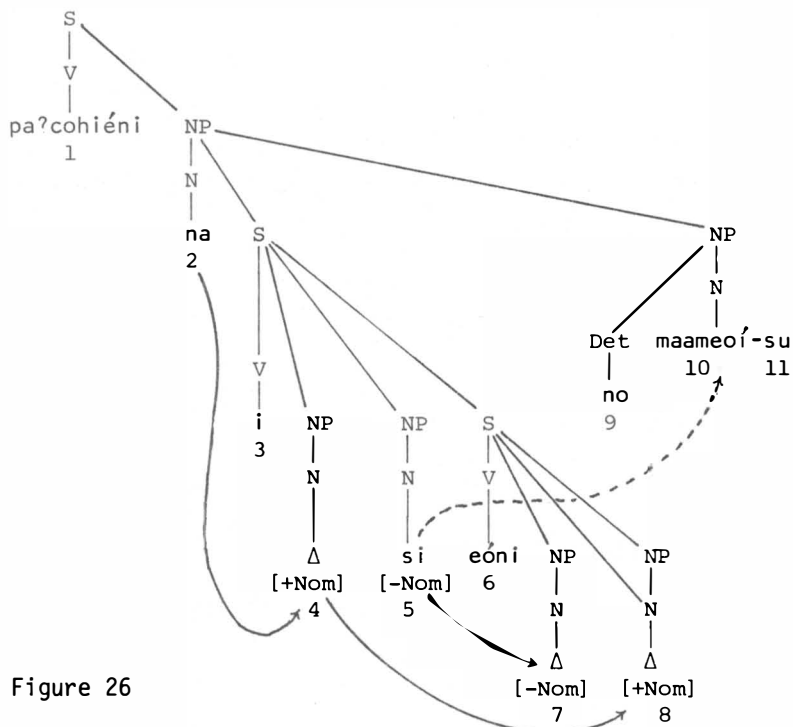


Figure 26

The curved lines indicate coreference; the [+Nom] Δ's represent the implied (location) subject of *eóni be located, occupy*, and the [-Nom] si and Δ designate the occupier, that is, the parents.¹³ As usual, the implied but missing element of a relative clause (always the subject in Tsou) is coreferential with the head noun of the NP, in this case na.¹⁴ Thus na is a kind of pronoun interpreted as the thing being located, the whereabouts being requested. The no maameoísu constituent is a possessive attribute of the head na. Since na refers to the whereabouts, no maameoísu specifies whose whereabouts it is that are under consideration. Marking actors grammatically as possessors is of course common in nominalisations in many languages, including English. Further examples:

- (75) *the boys covered up with the pot;* Tung 1964:434-435
 1 8 2-4- -4 5- -5 6
the ones which were covered by it, the pot which were small ones
 1- -1 4- -4 5- 3 -5 6 7- 7 8- -8

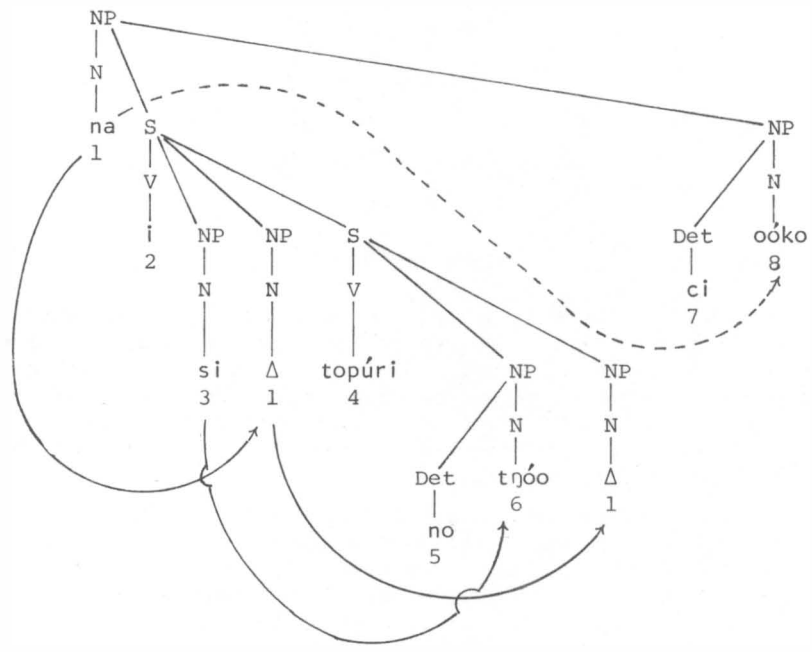


Figure 27

In this structure, ci is the appositional determiner, in effect a Philippine-like ligature, so that oóko is interpreted as in apposition to na rather than as a restrictive modifier of it.

- (76) *the boys covered up with the pot;* Tung 1964:434-435
 1 8 2-4- -4 5- -5 6
the small ones which were the ones that were covered
 1 8- -8 7- -7- -7 4- -4
by it, the pot
 5- 3 -5 6

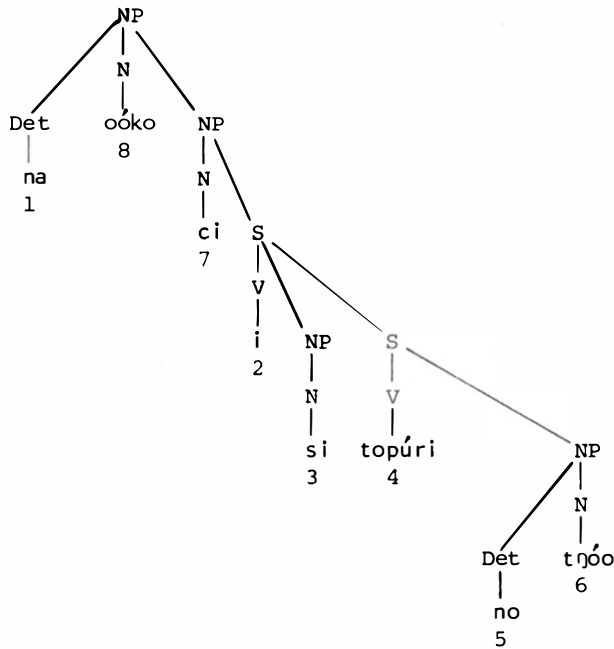


Figure 28

As can be seen from the tree diagrams, case marking 'particles' are thus not given a unitary analysis here; they are determiners when preceding lexical nouns, and nouns when preceding clauses. This can be taken to reflect their historical development: they must have originally been nouns – relator nouns or demonstrative pronouns – which served as derivational sources for the modern Tsou determiners while at the same time being retained as nouns in nominalisation constructions I am discussing here.

In effect, then, it almost appears as if Tsou has been a party to a conspiracy to eliminate all traces of the PAN denominal verbal focus affixes *-en, *ni-/in-, *-ana, and *iSi-. In order to accomplish this, it had to carry out two separate and unrelated processes. One of them was Aux-clutching, described above, which eliminated the verbal uses of these affixes by dropping all verbal affixes which were limited to main clauses. The second was the extension of a preexisting relator-noun relative clause nominalising strategy to the extent that it totally replaced lexical nominalisation involving this set of elements in their traditional positions in relative clauses, cleft equational sentences, and content interrogatives. These relator noun relative clause constructions are functionally equivalent to the deverbal noun constructions that SPR reconstructed for PAN, and so were able to replace these nouns in all positions, including cleft equationals.

Suspicious though it may seem, it is a fact that Tsou does have an auxiliary system and a nominalising strategy which are different in kind from those of the Paiwanic languages and different in degree from those of the Atayalic languages. At least the extension of these systems would seem to be an innovation in Tsou, and taken together, they have the effect of carrying out this cover-up.

The first innovation can be plausibly accounted for in terms of the Aux-clutching hypothesis proposed in this paper, but unfortunately I have no very good explanation to offer as to why the second substitution took place. In verbal main clauses, the originally denominal focus affixes were lost because of Aux-clutching, which was in turn motivated by the need for the presence of an auxiliary verb to express aspect and pronominal actors.

However, there seems to be no reason why this requirement would have affected equational sentences. That is, even if they underwent Aux-clutching when other intransitive clauses did, this shouldn't have resulted in the loss of the original derived noun forms, since equational clauses (unlike verbal clauses) presumably did not have distinct configurations when occurring in subordinate position. Thus it seems that they should have been retained in their original form in normal predicative and in relative clause constructions. The fact that such forms are not reflected in Tsou, then, leaves us with two questions to answer:

1. Why did the relator noun relative strategy get extended at the expense of the *-en, *ni-/-in-, *-ana, and *iSi- affixes in their nominalising functions, and
2. if this affix set originally functioned to derive nouns from verbs in pre-Tsou, why have no lexicalised fossils of the old nominalising process survived, as they have in various Oceanic languages under similar circumstances?

We might try resolving these difficulties by returning to the Pawley-Reid hypothesis (Pawley and Reid 1979) that *-en etc. did not begin as nominal affixes, but rather functioned originally as passive markers. Their loss would then be completely accounted for by Aux-clutching, and the absence of fossils would be accounted for. But then

1. how do we account for the specialisation of these forms to finite clauses in the first place? SPR does this in terms of the reanalysis of nominalised constructions, but we would be abandoning that explanation. And
2. what about the evidence from subgrouping? Once again, we are brought back to the paradox originally outlined in section 2.0: if the focus affixes were purely verbal in PAN, then either all the non-Tsouic languages form a subgroup which jointly innovated the nominal uses for these affixes, or there were independent parallel innovations in Atayalic, Paiwanic, Southern Tsouic, and possibly Rukai.¹⁵

5. CONCLUSION

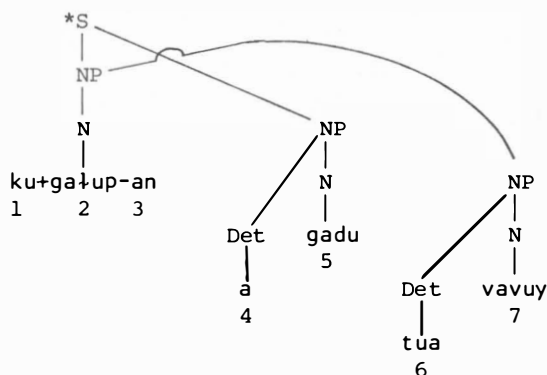
Aux-clutching then provides an explanation for the loss of denominal verbal focus affixes in Tsou, but there is no good reason why these same affixes in their nominal functions should have been lost without a trace.¹⁶ Unless such traces can be found, it may yet become necessary to seriously consider the radical hypothesis that there are only two primary subgroups among Formosan languages, Tsou and all the rest, as shown in Figure 4 above. The similarities between Tsou on the one hand and Rukai, Kanakanavu, and Saaroa on the other might then be accounted for in terms of areal influence and loans, which would help to account for the totally unrealistic phonemic inventory Tsuchida reconstructs for the supposed common parent of these four languages. That would mean that the PAN system reconstructed in SPR really applied to Proto-non-Tsou, that the nominal constructions as assumed in the earlier Pawley-Reid, Wolff, and Foley

reconstructions, and that Tsou itself reflects an earlier stage in which the 'focus affixes' being considered in this paper marked only finite verbs, and in which relative clauses were verbal clauses serving as attributes to relator nouns. In spite of the obvious problems with this approach, I think that future comparative lexical studies of Tsouic and Rukai will have to give it some serious attention.

NOTES

1. It is possible that the -n- of the Instrumental/Benefactive focus suffix -(n)eni is related to *-en. This affix is a Tsou innovation, and could be the result of the fusion of the Object Focus *-en with a following proximate demonstrative eni corresponding to the non-subject Agent. Alternatively, this -n- could be related to the genitive *ni, which is otherwise unattested in Tsou, again fused with eni.
2. It has been suggested that non-actor Focus constructions in such languages as Paiwan are *still* nominalised (Ferrell 1971:8), but this conclusion is based purely on the fact that such constructions mark agents with the Genitive case marker (cf. Egerod 1966:346 re Atayal). This reasoning however breaks down (for Paiwan at least) when confronted with facts about word order, since a sentence such as the following (Ferrell 1972:121)

ku+gaʔup-an a gadu tua vavuy
 1 2 3 4 5 6 7
I hunt boar on the mountain.
 1 2 7 3 4 5



in which the Nominative actant a gadu *the mountain* intervenes between the head word of the predicate and some other nuclear actants cannot be given a binary nominalised equational analysis.

3. cf. Young 1983 and references cited therein for a similar hypothesis regarding Proto-Indo-European.

4. Raleigh Ferrell (1972:126-127) gives examples of Tsou sentences which lack initial aspect auxiliaries and which contain clitic pronouns attached to main verbs. He notes that the 'preverbs have been stripped away for simplicity', a process which seems to have entailed his reattaching the associated clitic pronouns in the position in which one might otherwise expect to find them in other Formosan languages. I have not myself encountered any such sentences in my field work, and both Tung (1964:89) and Tsuchida (1976:97) explicitly deny that such constructions are possible in Tsou.
5. Note that this is not typologically anomalous, since what seems to be the same system appears in Tongan, Samoan and probably Mae and East Uvean (cf. Clark 1973:590). Mulder and Schwartz (1981:242) provide similar examples from Achenese, where the clitic pronoun refers to the actor in both an active sentence and its passive counterpart:

Drɔn nɪ̄ -pajoh bɔh -mamplam
 you 2sg eat fruit mango
 You eat the mango.

Bɔh -mamplam nɪ̄ -pajoh le-drɔn
 fruit mango 2sg eat by you
 The mango is eaten by you.

This looks like a good candidate for an aux-axing analysis operating on a basic Tsou-like clitic system.

6. The active non-focus forms *mo*, *mi*, *moso* (*miso*), *moh*, and *mio* (Tung's 'm-beginners'; Tung 1964:88) have an initial *m-*, which surely reflects PAN actor Focus **mu-*, while their passive counterparts, *i*, *o*, *os* (*is*), and (Tung's 'minus-m beginners'; Tung 1964:89), lack this element. The contrast between active and passive auxiliary verbs also appears elsewhere, as for example in the Atayal active perfective modal *nial*, derived from *uaial go*, e.g. (Egerod 1966:352):

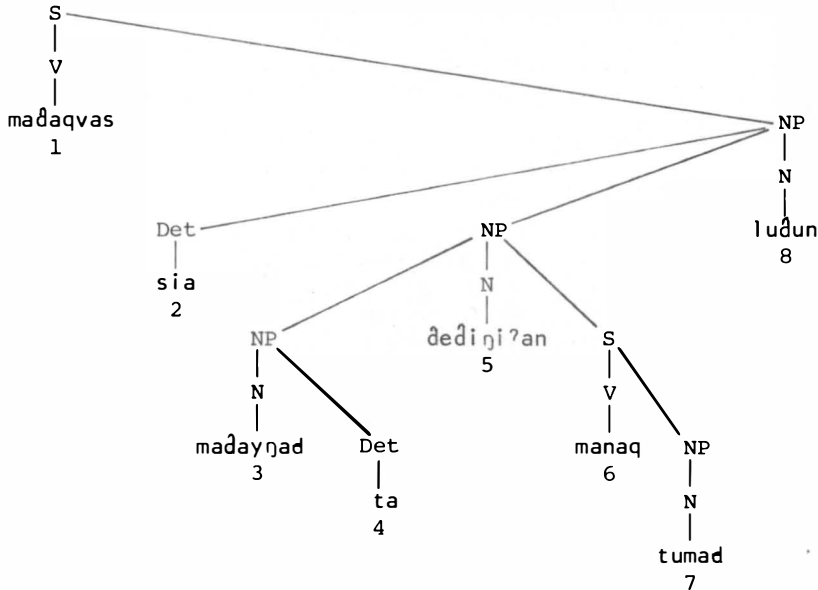
squ₁liq₂ nial₃ muah₄ km₄ut₄ itan₅
 1 2 3 4 5
 People have come to kill us.
 1 2 3 4 5

and the Amis denominal instrumental verbs, which can only appear in the passive (Chen 1985:6.2.3).

I think the Tsou non-future auxiliaries are related to the verb *go* which appears in various Formosan languages in auxiliary verb and motion verb uses. cf. Atayal *mosa?* ~ *usa?* *future action or event*, *ual* *past action or event* vs. *usa?* *to go* (Egerod 1965:271), Seediq *waada gone; perfective*, Saaroa *mia-* *to pass by* (Tsuchida 1976:76), etc. cf. also Palauan *mo go*, *be going to*, *m₁la*, *m₁le* *past* (Josephs 1975:129-131, 174-175, 272-275). A form *ta* is used in Amis (Chen 1982:281), Saisiyat, and Rukai with the same motion verb and auxiliary functions, especially in first person inclusive imperatives, and this is presumably related to the future auxiliary *te* (Tsou) and *tia* (Kanakanavu) and similar forms in some Oceanic languages.

7. 'Actor' can be defined in the lexibase framework being employed in this paper as the grammatical Patient of an intransitive clause or the Agent of a transitive clause. There is no single lexibase case relation to label this concept, which is also central to Fillmore's Subject Choice Hierarchy as well as to the distinction between accusative and ergative syntax. This notion is partly reflected in the man on the street's understanding of the word 'agent' and in the notions 'logical subject' in Chomskyan grammar and 'initial I' in relational grammar which are based on this same pre-theoretical concept.
8. It is interesting, though, that in Paiwan (Ferrell 1972:121-122), passives (my term, not Ferrell's; cf. Starosta 1974:363-364) exhibit a construction that could be seen as the result of a similar innovation: AF verbs apparently take only independent nominative pronouns, but passives all begin with a proclitic pronoun preceding a denominally affixed main verb. Based on Ferrell's examples, it seems that these clitics are always coreferential with the actor of the sentence, and thus could be seen as the result of 'Aux-Axing' (Starosta, Pawley and Reid 1981) applying in passive sentences to a Tsou-like clitic system.
9. If the denominal affixes in PAN were limited to transitive verbs in main clauses, then this requirement may also have applied to Locative Focus motion verbs such as *to go* as well as ordinary transitive verbs, if LF verbs were derived transitives as proposed in SPR.
10. In all the subsequent examples, the trees and bracketings are mine. I have also modified some of Tung's and Tsuchida's glosses in the Tsou examples to more directly reflect the syntactic structures, and I have replaced all instances of Tsuchida's /ə/ in the Tsou examples with Tung's more appropriate /ʔ/. Stress marks have mostly been omitted, since it is essentially phonologically predictable.
11. The development of relator relative clauses in Tsou is not totally unprecedented in Formosan languages. In Bunun, for example, we find that to compensate for the absence of a productive Locative Focus form, a deverbal noun *dediŋiʔan* 'place of happening' can be used to head a noun phrase containing a relative clause in non-Locative focus, e.g.

madaqvas sia maʔayŋad ta dediŋiʔan manaq tumad ludun [B35.1]
 1 2 3 4 5 6 7 8
The mountain where the man shot the bear is very high;
The mountain, the man's place to shoot the bear, is high.
 2 8 4 3 5 6 7 1



12. I think that the yang relative clauses of Bahasa Indonesia are probably to be analysed in a similar fashion.
13. The coreference should proceed from clause to clause, I think, if the Sisterhead Constraint is to be maintained in grammatically conditioned coreference relationships. That means that we must assume that the auxiliary clause, whose only overt actant is an accusative Agent, should be considered to also imply a dummy Nominative coreferencing both the Patient of the lower clause and the na head of the next higher construction.
14. One difficulty with this analysis concerns constructions such as Tsuchida's interrogative examples, given as examples 67 and 68 above and repeated here:

(67) NP [cuma] NP [na+ S [+ko fr+ʔs-a]] Tsuchida 1976:102
 1 2 3 4 5
What is the one which was covered by you?
 1 2- -2- -2 5 4 3

(68) NP [cuma] NP [na+ S [+ko fr+ʔs-eni]] Tsuchida 1976:103
 1 2 3 4 5
What is the one which was covered with by you?
 1 2- -2- -2 5- 4 -5 3
What did you cover with? What did you use for a cover?

If na+ in this sentence is analysed as a Determiner, what is it doing being followed by a clitic pronoun ko+? If na+ is a noun, of course, this is not in itself strange, since it is normal to have nouns followed by possessive clitic pronouns. Unfortunately, though, ko+ is the wrong kind of clitic: it occurs only with four of the sixteen auxiliary verb forms distinguished by Tung (1964:109-110), mi, os, te and tena, and never occurs with nouns otherwise to mark possession. These examples are the

- only ones of this kind I have found in Tsou, though, and are more likely the result of non-systematic elision of the auxiliary. Since the auxiliary is partly recoverable from the form of the clitic, its loss in this environment does not result in the loss of very much information.
15. Actually the situation in Rukai is quite similar to that in Tsou (cf. Li 1973). It will not be possible within the limits of this paper, though, to follow up the synchronic and diachronic implications of this fact.
 16. It has been suggested by Otto Chr. Dahl (personal communication) that I may simply be making too much of the noun-verb distinction, which as we know is frequently difficult to justify in certain constructions in many Austronesian languages. He points out (quite correctly, in my opinion) that such a distinction cannot be based purely on semantics, or even (in Malagasy, at least) on cooccurrence with pronominal suffixes. However, if we were to abolish it and lump the two categories into a single class, say, 'Contentive', it would still be necessary in an explicit (generative) grammar to account for the range of constructions in which members of this class occur. Supposing this could be done successfully without distinguishing nouns from verbs, we would still need to account for the different inventories of constructions in Tsou and the languages that are supposed to subgroup with it. I think that the end result would still be the same dilemma encountered with separate (though overlapping) classes of nouns and verbs.

BIBLIOGRAPHY

ASAI, Erin

- 1953 *The Sedik language of Formosa*. Kanazawa: Cercle linguistique de Kanazawa. Kanazawa University.

BAKER, Carl L.

- 1978 *Introduction to generative-transformational syntax*. Englewood Cliffs: Prentice-Hall.

BRUCE, Les

- forth- Serialisation: the interface of syntax and lexicon. *Papers in coming New Guinea linguistics* No. 24. PL, A-70.

CHEN, Teresa M.

- 1985 *Verbal constructions and verbal classification in Nataoran-Amis*. (Ph.D. dissertation, University of Hawaii, 1982; pagination from from the published version.) PL, C-85.

CLARK, Ross

- 1973 Transitivity and case in Eastern Oceanic languages. *Papers of the First International Conference on Comparative Austronesian Linguistics I: Oceanic*. *Oceanic Linguistics* 12/1-2:559-606.

- EGEROD, Søren
- 1965 Verb inflection in Atayal. *Lingua* 15:251-282.
- 1966 Word order and word classes in Atayal. *Language* 42:346-369.
- FERRELL, Raleigh
- 1971 Ergative, passive or focus-marking? Elicitation of verb syntax in Indonesian languages. *Colloques Internationaux du Centre National de la Recherche Scientifique, Sciences Humaines, Methodes d'Enquête et de Description des Langues sans Tradition Écrite*. 9pp.
- 1972 Verb systems in Formosan languages. In Jaqueline M.D. Thomas and Lucien Bernot, eds *Langues et techniques, nature et société*, I. 121-128. Paris: Klincksieck.
- FOLEY, William
- 1976 Comparative syntax in Austronesian. Ph.D. dissertation, University of California, Berkeley.
- HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds
- 1982 *Papers from the Third International Conference on Austronesian Linguistics*. vol.2: *Tracking the travellers*. PL, C-75.
- HARVEY, Mark David
- 1979 Subgroups in Austronesian. BA Honours subthesis, the Australian National University.
- HORN, George M.
- 1975 On the nonsentential nature of the POSS-ING construction. *Linguistic Analysis* 1/4:333-387.
- JOSEPHS, Lewis S.
- 1975 *Palauan reference grammar*. [PALI reference texts: Micronesia.] Honolulu: University Press of Hawaii.
- LI, Paul Jen-kuei
- 1973 *Rukai structure*. (Academia Sinica special publications no. 64.) Taipei: Academia Sinica.
- MEI, Kuang
- 1972 Pronouns and verb inflection in Kanakanavu. *Tsing Hua Journal of Chinese Studies* 14/1-2:207-232.
- MULDER, Jean and Arthur SCHWARTZ
- 1981 On the subject of advancements in the Philippine languages. *Studies in Language* 5/2:227-268.
- NAYLOR, Paz B., ed.
- 1979 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*. [Michigan Papers on South and Southeast Asia no.15] Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

PAWLEY, Andrew K., and Lawrence A. REID

- 1979 The evolution of transitive constructions in Austronesian. In Naylor, ed. 1979:103-130.

REID, Lawrence A.

- 1982 The demise of Proto-Philippines. In Halim, Carrington and Wurm, eds 1982:201-216.

ROSS, John Robert

- 1967 Auxiliaries as main verbs. In W. Todd, ed. *Studies in philosophical linguistics*, Series one, pp.77-102. Carbondale, Illinois: Great Expectations Press.

SCHACHTER, Paul

- 1976 A non-transformational account of gerundive nominals in English. *Linguistic Inquiry* 7/2:205-241.

STAROSTA, Stanley

- 1974 Causative verbs in Formosan languages. *Papers of the First International Conference on Comparative Austronesian Linguistics II: Proto-Austronesian and Western Austronesian. Oceanic Linguistics* 13/1-2:279-370.
- 1977 Affix hobbling. *Working Papers in Linguistics, University of Hawaii* 9/1:61-158.
- 1982 Mandarin Chinese case marking: a localistic lexicase analysis. *Working Papers in Linguistics, University of Hawaii* 14/3:

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

- 1982 The evolution of focus in Austronesian. In Halim, Carrington and Wurm, eds 1982:145-170.

TSUCHIDA, Shigeru

- 1976 *Reconstruction of Proto-Tsouic phonology*. (Monograph series no.5, Study of languages and cultures of Asia and Africa.) Institute for the study of languages and cultures of Asia and Africa. Tokyo: Tokyo Gaikokugo Daigaku.

TUNG, T'ung-ho

- 1964 *A descriptive study of the Tsou language, Formosa*. (Institute of History and Philology special publication no.48.) Taipei: Academia Sinica.

WOLFF, John U.

- 1973 Verbal inflection in Proto-Austronesian. In Andrew B. Gonzalez, ed. *Parangal kay Cecilio Lopez* (Essays in honor of Cecilio Lopez on his seventy-fifth birthday), 71-94. (*Philippine Journal of Linguistics special monograph issue no.4*.) Quezon City: Linguistic Society of the Philippines.
- 1979 Verbal morphology and verbal sentences in Proto-Austronesian. In Naylor, ed. 1979:153-168.

YOUNG, Susanne J.

- 1983 The medio-passive r-forms in Indo-European. Ph.D. dissertation, University of Hawaii at Manoa.

THE ROLE OF COMPUTERS IN DICTIONARY-MAKING AT THE UNIVERSITY OF HAWAII

Robert Hsu

1. DICTIONARIES vs. LEXICAL FILES

In a typical dictionary the different types of information in an entry are indicated by a hodge-podge of different conventions such as fonts, abbreviations, brackets, and so forth. Fig. 1 shows an excerpt from a conventional dictionary (of Woleaian). Headwords are shown in boldface, base forms in small caps, etc. For working in the computer, however, we use a representation which I will call a 'lexical file', in which such functions are overtly labelled in some uniform manner. We use short mnemonic abbreviations at the left of the line, and begin each new type of information on a new line, as in Fig. 2. In this example, 'hw' labels headwords; 'ba', base form; 'df', definition; and so on.

We call each such labelled type of information a 'band'. An entry in a dictionary corresponds in the lexical file to a sequence of such bands. A period or full stop before a band name marks the beginning of an entry. There is no limit to the number of bands one may invent for a lexical file. Typically there would be 20 to 30 bands; one of our files has over 200. Substructure within an entry is indicated by a system of numbers preceding these labels which we will not describe here. There is also a system for indicating subentries and sub-subentries, etc.

These conventions are by no means the only ones one could use for encoding lexical files, and they certainly do not handle everything one might wish. They do however have the virtue of being quite easy to work with. We have on the whole been satisfied with them in the course of working with over thirty languages.

Overtly labelling the information and its structure, rather than stringing it together into continuous paragraphs as in a dictionary, makes it easy for computer programs to identify and manipulate the various types of information. This is the basic reason for working, in the computer, with lexical files even though the principal goal may be the production of a dictionary.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 313-328.
Pacific Linguistics, C-88, 1985.

© Robert Hsu

2. LEXICAL FILE AS A REPOSITORY

The advantages of working with a lexical file also affect any dictionary that might be produced from the file. These advantages fall into several categories. One is a consequence of treating the file as an active and growing repository of information rather than primarily as a manuscript to be edited and published. Having a structure such as a system of bands, and having the ability to invent new bands whenever needed, facilitate the organisation and filing of information as it is gathered. This facilitation effect is often noticed by lexicographers as soon as they begin putting information into a lexical file, even before any computer processing has taken place.

Another effect of treating the file as a repository is that one is free to include information that might not be appropriate for a published dictionary. One might have a band, for instance, containing coding of semantic domains, or one for references to field notebook pages or to tape counter numbers or names of speakers who supplied the form. There may be a band for private comments such as "check this form again", or "listen to the last vowel again". There may be bands for comparative data from related languages. Fig. 3 shows an excerpt from a lexical file, printed in a more readable format than Fig. 2, in which we see some bands (e.g. CO, RF, and the various comparative bands) that are probably not meant for publication in a dictionary.

Yet another result of treating the file as a repository is that, since one is freed (at least while gathering the information) from the economic, political, and other considerations of what should actually be published, one can better focus on the gathering and verifying of the linguistic and cultural materials in the file.

3. EASE OF MAINTENANCE OF THE FILE

A second category of benefits arises from the ease of maintenance of a computer file. Being in computer storage, the file can be readily edited at a terminal as new and corrected information is gathered. The band format that we use is reasonably easy to work with in this respect. After editing, a clean copy of the updated file can easily be printed out. The format of the printout can be altered to suit the purposes of the user. The typical format we use for working copy is illustrated in Fig. 4, which corresponds to a part of what appears in Fig. 2. The two boxes correspond to the boxes in Fig. 2 and Fig. 1. The headword band is given prominence by a considerable overhang at the left, entries are set off by blank lines, subentries are indented, page numbers are provided, etc. On the whole, this format is much easier to read than the original file itself, and provides more space for hand-written notes. (For reference copies we use a more compact printout format.) Other formats can easily be devised.

It is of course easy to print out copies for distribution to co-workers. It is also a simple matter to make copies of the file on computer tape, to share with colleagues at other institutions — provided that their computers are compatible. In these respects computer files are much preferable to slip files or notebooks.

4. DEVELOPING THE FILE

A third class of benefits arises from the ability of the computer to help in refining and expanding the file itself. The computer can be instructed to make changes to the file. It can also be instructed to run checks on the file. An example of automatic changes is alphabetisation. New words can be sorted into their proper alphabetical places; and if a new alphabetical order is established, the entire file can be automatically re-alphabetised according to the new system. Also, several dictionaries of the same language can be sorted together to develop a new and updated dictionary. Fig. 5 shows a consolidation, a sorting together, of several dictionaries of Tahitian dating from the 19th century which is being augmented by contemporary lexical data and made into a historical dictionary by Jack Ward. The five entries in the box come from five different dictionaries, as indicated by the names of the headword bands. Ward will now conflate such groups of entries, adding notes and material of his own.

Another example of automatic changes is the task of respelling all the words in a lexical file according to new spelling conventions. The Marshallese words in the Marshallese lexical file once had to be respelled by the application of some 50 ordered context-sensitive rules. Since this had to be done to not only the headwords but also to the thousands of Marshallese words in the illustrative sentences, this job could have been accomplished economically only by computer. A similar respelling, though very much simpler, is currently being undertaken for the Chamorro file.

Automatic checks also can be run by the computer on a lexical file. For instance, a program can check that all cross-references in the file refer to actual headwords in the file, or that all words used in illustrative sentences also appear as headwords. A program can also check spelling to the extent of reporting occurrences of prohibited or rare letter sequences.

Human checking of material in the file can be facilitated by computer generated listings: the computer can print out all headwords containing certain specified phonological sequences for further checking with native speakers; or if fish names or plant names, say, have been entered in special bands, the computer can print out all entries having such bands, and the printout can be sent to the field or to a specialist for checking and expansion. Fig. 6 shows a portion of what we call a bandsort, a listing of bands (here DF and GR) extracted from their entries and regrouped by band name rather than by entry. This has proved to be, among other things, a very effective proofreading aid, as it allows one to scan all instances of one particular band without being distracted by other bands. Other printouts can also be made which bring together similar phrases occurring in, for instance, the definitions, so that one can easily catch lack of parallelism in phrasing.

5. USES OF THE FILE

So far we have discussed how the computer can help the lexicographer maintain and develop the file itself. Now we ask, of what use is such a continually refined and expanded computer file? One of the most frequently voiced reasons for committing a lexical file to the computer is the need for easily making an English index, called a finderlist. In the sample printouts of lexical files you will have noticed asterisks attached to certain English words in the definitions. These were placed there by the lexicographer to mark words that are to be extracted by the program which makes a finderlist.

Fig. 7 is an example of such a computer-generated finderlist. The box surrounds a phrase extracted from the definition band in the first boxed entry in the previous Woleaian examples (Fig. 4). The asterisk on the word 'tuna-fish' was the signal to the program to extract the entire phrase and place it under the keyword 'tuna-fish' in the finderlist. Our programs recognise a large number of such conventions that the lexicographer can use in definition bands to delimit keywords and phrases. In this way a relatively complex finderlist can be generated automatically from a lexical file which has been appropriately decorated with stars and other special symbols.

Another possibility, of interest to the comparative linguist, is that of combining the finderlists of several related or neighbouring languages. Fig. 8 shows a page from such a combined finderlist generated from the lexical files of 12 languages of Micronesia. Even though the 12 files were of uneven coverage, it has nevertheless served as an important source of data for a project in comparative reconstruction.

Possibly the most significant use of the file in the long term is as a reference work that can be consulted by means of appropriate computer programs. Since the different types of lexical information are explicitly labelled, programs can be written to extract entries, or parts of entries, containing specified information, as aids to the study of the language or the culture. Fig. 9 shows portions of four printouts of Palauan nouns classified according to the vowel of the third person possessive suffix. These printouts were made from the Palauan lexical file for a study of the distribution of these vowels. Fig. 10 is a portion of an index based on codes for semantic domains entered in a Kiribatese (Gilbertese) dictionary. This excerpt shows words having to do with food (FOO) and with geology and geography (GEO). Since the particular concerns of a culture tend to be reflected in its vocabulary, printouts such as this can be used as aids in studying and appreciating a culture.

A final example of uses of a lexical file is the making of a dictionary from the material in the file. At any time the whole file or any subset of it can be extracted, to make a reference dictionary, or a briefer glossary for school use, or a gazetteer of place names or fish names, or any other such list. As for printing, reproduction of an appropriately formatted computer printout might be adequate for some purposes, but regular letterpress quality printing is also possible through computer-driven photocomposition. Since the material is in computer-readable and structured form, it can be reformatted by program for input to a photocomposition machine. The output of the latter is a photographic master suitable for photo-offset duplication. This route by-passes the need for retyping the entire text (which would entail another proof-reading step) and is in most cases less expensive. For concreteness I have included a sample of what the photocomposition tape looks like (Fig. 11). This is, in fact, a portion of the tape used to photocompose the Woleaian dictionary. The portions enclosed in boxes again correspond to the two entries we have been following. This is, of course, not intended for human consumption. Embedded in the text are codes for shifting to italics, bold, roman, etc., for starting a new paragraph, and for other typographical functions.

After the dictionary is published the lexical file need not go into retirement. It remains a valuable resource, a database for further research, and is available for further development and use. Additional editions of the published dictionary can be produced as the file grows.

6. THE ROLE OF THE COMPUTER IN DICTIONARY-MAKING

The principal use of the computer in dictionary-making then, at least at the University of Hawaii, is in helping to maintain and develop the lexical file, a resource of many uses, one of which is the production of a dictionary.

The computer has at least two other possible uses in dictionary-making, not directly related to lexical files, which I will briefly touch on in conclusion. They involve the computer as an aid in finding and defining words in the language. The first method was first used by Vern Carroll for Nukuoro. He called it generative elicitation. We have since used it many times. It consists in having the computer produce all possible forms of words according to the known phonotactics of a language, Fig. 12 is a part of such a printout of trisyllables for Motu. A native speaker is invited to read through such a list and to note all forms which actually occur in the language. In the case of Nukuoro, Carroll further had the computer generate all morphotactically possible derivations of the roots discovered from the first list. In either case, such a project is not to be embarked upon lightly, since an exhaustive printout of all disyllables, not to mention trisyllables, for a language with even a small inventory of consonants and vowels runs into the tens of thousands, and the rate of return is typically very low. It also requires literate and very patient native speakers.

The other method of finding words is the quite common one of making concordances from text. Fig. 13 is a sample of an interlinear concordance of an American Indian language of the Pacific North-West. A concordance not only finds all the words or morphemes in a text but also brings together all the contexts of a given morpheme or word, so that it can be studied in all its uses and meanings in the text.

These two additional uses of the computer complete our quick survey of the computer's role in dictionary-making.

	Woleaian-English
tafey	
tafey (tafeyə). 1. N medicine. <i>Yoor t. lan sipitaal</i> There is medicine in the hospital. 2. VI to be treated. <i>Ye sa t. pesheei</i> My leg has been treated.	
tafeya (tafeyaa). VT give medicine to him, treat him. <i>Tagota u a ye t. sar uə</i> . The doctor treated the child. <i>Ye t. Mary</i> He gave medicine to Mary.	
tafiy (tafi-a). VT open it with fingers (referring to vagina), pull it apart. <i>T. lag!</i> Open it (the vagina) up!	
tafish (tafishh). 1. N trap, snare. <i>Ye lag seuw t. shiul gashi la yaai</i> . There is a trap on my tuba tree. 2. VN to trap, snare.	
tafishfish (tafishfishh). VI to sparkle. <i>Ye t. lag yaf uə reel timu uə se bbul</i> . The burning house gives away sparkling fire.	
tafishhiy (tafishhi-a). VT trap it, set a trap for it, snare it. <i>T. gesh uə!</i> Set a trap for the rat!	
tafitē ₁ (tafitaf). VN to open (vagina), pull apart. <i>Mwal uə ye tau t. tingiy</i> . The man habitually opens (vagina).	
tafitē ₂ (tafitaf). N fishing kit. <i>Ifa t. uə yaai?</i> Where is my fishing kit?	
tafusiufius (tafusiufiusiu). VI to be spotted, have a small mark. <i>Ye t. pesheel</i> . His legs contain many marks. <i>Ye t. siilo uə</i> The pig is spotted.	
tag ₁ (tagē). VI to float in shallow water, sail in shallow water. <i>Ye t. uə uə</i> The canoe is able to sail in the shallow water.	
tag ₂ (tagē). DIR upward, eastward, up. <i>Ye sa rig t. sar uə</i> The child ran eastward.	
tag ₃ (tagiu). N needle fish.	
-tagē (-tagē). [directional suffix] upward, up, eastward. <i>gabitageyei</i> , make me go up. CF. tag ₂ .	
tageey (tagee-a). VT. ride it, sail in it. <i>Re t. wa uə</i> . They sailed on the canoe.	
tageloa (tage-loa). VI to surf. <i>Re t. Waikiki</i> They are surfing in Waikiki.	
tageshaliyal (tageshali-yalə). N sunrise. SYN tegaliyal.	
tagiulwosh (tagiuli-woshə). N a kind of needle fish.	
tagiunal (tagiuli-lalə). N trumpet fish.	
tagiur ₁ (tagiuriu). VI to face, turn. <i>Ye t. tanqiyei</i> She faced away from me.	
tagiur ₂ (tagiuriu). N back (anatomical).	
tagiuraar (tagiuranaar). N sword fish.	
tagiuriu (tagiuriu-a). N my back. <i>Ye biuu shiul t.</i> My backbone is broken.	
tagiuriupaai (tagiuriu-paai). N a kind of lizard. <i>Resepaal re muommuwongo t.</i> Japanese people ate lizards.	
tagiuter (tagiuteriu). N a kind of needle fish. <i>seuw t. semal t.</i> , a needle fish.	
tagiyat (tagiyata). VI. ADJ. (to be) high, tall, great, lofty. <i>Ye t. wa uə</i> . The canoe is high.	
tagiyetaat (tagiyataatə). N high place, raised place. <i>Ye matt wool seuw t.</i> He is sitting on a raised place.	
tagomeliuw (tagomeliuwa). N a kind of breadfruit with smooth surface and white flesh. <i>Ye iyeri sefash t.</i> He picked breadfruits from a tagomeliuw tree	
tagomwaailey (tagomwaaileyali). VI to be dizzy, go around in a circle. <i>Ye sa t. sar uə</i> . The child is dizzy as a result of going around in a circle.	
tagulugul (tagulugulu). VI to spin (many times), rotate, turn. <i>Ye t. misiil uə</i> . The machine is turning.	
tagun (tagunu). VI to turn, return, shift (of wind). <i>Ye t. tangi</i> . He turned away from her.	
tagutog (tagu-tagu). VN to chase, block. <i>Re t. ig uə</i> . They are chasing fish. CF. taguuw.	
taguuw (taguu-a). VT chase it, block it. <i>Re t. ig uə</i> . They chased the fish. CF. tagutog.	
taguw (taguu-a). N yellow-fin tuna-fish.	
tai (taə). ASP [negative] not. <i>Ye tai goang</i> . It is not me. <i>Ye tai lag</i> . He did not go.	
taig (taigo). [rang in Faraulep dialect] 1. N turmeric. <i>Ye lag tingar t. me reel melewe sin</i> . She went to ask for turmeric from her mother. 2. VI. to apply turmeric on one's body. <i>Re t. sar kawe</i> . Those children are putting turmeric on their bodies.	
taiif (taiifa). N platform of a house, stones around the house used to keep the gravel from spreading out. <i>Ye matt wetaiif</i> . He is sitting on the platform.	
taii (taiiə). N mountain, hill. <i>Ye toulap t. uoal Hawaii</i> There are many mountains in Hawaii.	
taikeil (tai-kaila). VI. ADJ. (to be) weak, unhealthy. <i>mal t.</i> weak person. CF. kail	

Figure 1

.hw	tagun
ba	tagunu
ps	vi-
df	to *turn, *return, *shift (of wind)
sx	ye t. tangi he turned away from her
.hw	tagutog
ba	tagu-tagu
ps	vn-
df	to *chase, *block
sx	re t. ig they are chasing fish
see	taguuw
.hw	taguuw
ba	taguu-a
ps	vt-
df	*chase it, *block it
sx	re t. ig we they chased the fish
see	tagutog
.hw	taguw
ba	taguwa
ps	n-
df	yellow-fin *tuna-fish
sem	*fish-species
.hw	tai
ba	tai
ps	asp.
gr	negative
df	*not
sx	ye tai gaang jit is not me
ps	ye tai lag he did not go
.hw	taig
ba	taigo
al	rang [in Faraulep dialect
1ps	n-
1df	*turmeric
1sx	ye lag tingar t. me reel melewe sin she went to ask for turmeric from her mother
2ps	vi-
2df	to apply *turmeric on one's body
2sx	re t. sar kawe those children are putting turmeric on their bodies
.hw	taiif
ba	taiifa
ps	n-
df	*platform of a house, stones around the house used to keep the gravel from spreading out
sx	ye matt wetaiif he is sitting on the platform
sem	*house-parts: platform
.hw	taiit
ba	taiita
ps	n-
df	*mountain, *hill
sx	ye toulap t. uoal Hawaii there are many mountains in Hawaii
.hw	taikeil
ba	tai-kaila
ps	vi-, adj.
df	(to be) *weak, *unhealthy
px	mal t. weak person

Figure 2

- 408 .HW I 52
 GC P
 GG *DIRECTIONAL PREP.
 GL *FROM
 PI *LONG
 EG NAU E AU I #HAATAE: I CAME FROM #HAWAII
- 409 .HW I 53
 GC P
 GG *DISTRIBUTIVE PREPOSITION
 GL *AMONG, *BETWEEN
 PH SE TAMA I TAATOU: ONE PERSON AMONG US
 CO CHECK GLOSS
- 410 .HW IA
 GC PN
 PPH *IA | HE, SHE + (B1)
 GR SEE GRAMMAR FOR DETAILS OF USAGE
 GG 3RD PERS. SG. PN.
 GL *HE, *SHE, *IT; *HIN, *HER, *IT
 PI *EM
 CP AAIA: USED AFTER #I, KI; AIA: USED AFTER SUFFIXED VERBS, ETC.
 PCO IA | HE, SHE + (IH)
 NAN IA | HE, SHE + (R1-28)
 NUK IA | X + (C3-248)
 SIK -IA | 3RD PER SG 'HE' (OBJ) + (S1-15)
- HJ NAIA
 GC
 GL *WITH *HIN (*HER, *IT)
 PI
 EG
- HJ SOROIA
 GC
 GL BY *HIM*SELF, *HERSELF, *ITSELF
 PI *EM TASOL, EM WAMPALA, EM YET
 LUA SO'IA | ALONE, BY HIMSELF + (S2-68)
- 411 .HW IAA-
 GC PREFIX
 GL *ALONG, *THROUGH
 PI *LONG
 CP AA-, VAA-, I
 CO GET BETTER GLOSS. #SIEKI SAYS #IAATAI IS 'ALONG THE EDGE OF THE LAGOON', #VAATAI IS 'LAGOON SIDE OF SOMETHING'
 BP (GR-029)
- 412 .HW IAA
 PS EXCLAM
 GL *EXCLAMATION MADE WHEN GIVING SOMETHING: *HERE!
 PI HAP *TOK YU HEKIB LONG TAIN YU GIVIN SAMING LONG WAPAPALA HAW

Figure 3

WOLEAI LEXICAL FILE	
.hw	taquw
	ba taquwa
	ps n.
	df yellow-fin *tuna-fish
	sem *fish-species
.hw	tai
	ba tai
	ps asp.
	qr negative
	df *not
	sx ve tai qaang it is not me
	sx ve tai laq he did not go
.hw	taiq
	ba taiqo
	al ranq in Faraulep dialect
	1ps n.
	1df *turmeric
	1sx ve laq tingar t. me reel meleve sin she went to ask for turmeric from her mother
	2ps vi.
	2df to apply *turmeric on one's body
	2sx re t. sar kave those children are putting turmeric on their bodies
.hw	taif
	ba taiifa
	ps n.
	df *platform of a house, stones around the house used to keep the gravel from spreading out
	sx ve watt wetaiif he is sitting on the platform
	sem *house-parts: platform
.hw	taiit
	ba taiita
	ps n.
	df *mountain, *hill
	sx ve toulap t. wool Hawaii there are many mountains in Hawaii
.hw	taikeil
	ba tai-kaila
	ps vl., adj.
	df (to be) *weak, *unhealthy
	px mal t. weak person
	see kail
.hw	taiko
	ba taikoo
	lo Jap.
	ps n.
	df *drum
	sx re sa qakiunqiu t. ve they have sounded the drum

Figure 4

SORTED FE, FI SECTIONS

```

11 .js feaa
    pos vn
    fr *penser, *hesiter, *douter

12 .an fea'a
    enq *doubt, *vacillate, *agitated
    cf fea |hov
    | wheanqaanqa

13 .sv fe'a'a
    1pos vi
    lenq to *coqitate, to *think, to *hesitate, to be *undecided
    lenq to *doubt

    2pos n
    2enq *doubt

..sv fea'a'ra'a
    enq *doubt, *hesitation

..sv fea'a-'ore
    enq *thoughtless, *unconcerned

..sv fea'a-piti
    enq to *hesitate between two alternatives

14 .da feaaore
    dr fe'a'a'ore
    pos a
    enq *thoughtless, *unconcerned

15 .da feaabiti
    dr fe'a'apiti
    pos v-D.
    enq to halt between-*two-opinions

16 .ma feaabiti
    pos v int
    fr *hesiter entre deux idees

17 .js feaa-piti
    pos vn
    fr *hesiter entre deux partis

18 .an fea'apiti
    enq a *halt between two *opinions, to *vacillate
    ma fea'a |doubt: |piti |two

19 .ln fea'a piti
    fr *hesiter entre deux idees
    ex 'ua fea'a piti to:'u sana'o *no: *te haere i No'orea |j'hesite (ma penseie hesite) a2 aller a
    No'orea

```

Figure 5

TYOA	29841	DF	WEDGE PLACED ACROSS THE HULL OF THE CANOE, *THWART
PEL	21769	DF	WHITE *COCONUT *MEAT
MOUS	18844	IDF	WHITE *HAIR, GRAY HAIR
GEFAAMASH	9371	DF	WHITE *MORNING-GLORY (XIPONCEA TUBA)}
UBIL	30005	DF	WHITE *SHELL, *COWRY, BIVALVE *MOLLUSC
RESH \$2	23151	DF	WHITE CAPPED *NODDY
REEHASOM	23015	DF	WHITE MATURE *TURTLE *EGG
GIYEGYI \$2	11335	DF	WHITE OR FAIRY *TERN
PFNGAGIL	21815	DF	WHITE PORTION OF SEA *CLAM
ITFIU	12627	DF	WHO?
GFPAABUTA	10270	DF	WISH *BAD-LUCK CN HIM
GFFAIYEBUTA	9430	DF	WISH HIM *BAD-LUCK
GFFAIYEFISHIYI	9442	DF	WISH HIM *GOOD-LUCK
GEPAAFISHIYI	10277	DF	WISH HIM *GOOD-LUCK
GAULESHIMW	6370	2DF	WITH THE HEAD BENT FORWARD
GEMASE	9902	DF	WITNESS HIS DEATH
GILIGIL \$1	11026	DF	WOMAN'S *CUTTER FOR BANANA OR LAVALAVA FIBERS
SIF	25878	DF	WOMAN'S *SKIRT, *GRASS-SKIRT
REWA	23167	DF	WOODEN NATIVE *COMP (MEN'S COMB ONLY)
PERAS \$1	21956	DF	WOVEN *COCONUT *LEAVES USED ON SAILING CANOES
GILIECHO	11059	DF	WOVEN *COCONUT *LEAVES FOR STORING DRY COPRA, *COPRA-*BASKET, *MAT FOR DRYING COPRA
TFPELIUNWAN \$2	28489	DF	WOVEN WHITE *COCONUT *LEAVES USED AS *LEIS
TIUGIUM	29118	IDF	WRAPPED PRESERVED *BREADFRUIT COOKED BY AN UNDERGROUND OVEN
MWE TAR \$1	19477	DF	YELLOW *FAT ON THE *INTESTINE
MOSHORANG	31824	DF	YELLOW *REEF, YELLOW *CORAL
JAGUW	27120	DF	YELLOW-FIN *TUNA-FISH
GURUB	17853	DF	YOUNG *COCONUT
WOLSHIG	31691	DF	YOUNG *COCONUT TREE
UB \$3	30000	DF	YOUNG *COCONUT, YOUNG *COCONUT MEAT
MWAG	18976	DF	YOUNG *NEEDLE-FISH
ROWAF	23547	DF	YOUNG *OCTOPUS
IL \$4	12170	DF	YOUNG *SHOOTS SURROUNDING AN OLD PLANT, *YOUNG-*PLANT
LIPEDMASHING	15834	DF	YOUNG *SKINNY PERSON
TAR \$5	27697	DF	*TARO, YOUNG TARO PLANTS SURROUNDING THE PARENT TARO PLANT, YOUNG *PLANT, YOUNG ONE
USHUGA	30281	DF	YOUNG BREADFRUIT
LAUMW	14957	DF	YOUR *CHILD
PAUMW	21634	DF	YOUR *HAND, YOUR *ARM
-IS	12580	GR	*NUMERAL-*CLASSIFIER FOR A CLUSTER OF BANANA HANDS; #APPEARS ONLY WITH XSE- *ONE
-UT	30324	GR	*NUMERAL-*CLASSIFIER FOR A HANDFUL OF SPONGE-LIKE OBJECTS (USUALLY BREADFRUIT); #APPEARS ONLY WITH XSE- *ONE
-BIS	1426	GR	*NUMERAL-*CLASSIFIER FOR A SMALL AMOUNT
-KIIT	13941	GR	*NUMERAL-*CLASSIFIER FOR A SMALL AMOUNT
-KIUT	14140	GR	*NUMERAL-*CLASSIFIER FOR A SMALL AMOUNT; #APPEARS ONLY AFTER XSE- *ONE
-NMWIT	18743	GR	*NUMERAL-*CLASSIFIER FOR A SMALL AMOUNT; #APPEARS ONLY WITH XSE- *ONE
-PAIU	21293	GR	*NUMERAL-*CLASSIFIER FOR AN ARM LENGTH; #APPEARS ONLY WITH XSE- *ONE; #FOR TWO ARM LENGTHS, XSENGAF IS USED
-MAL	17237	GR	*NUMERAL-*CLASSIFIER FOR ANIMATE BEINGS SUCH AS ANIMALS, HUMAN BEINGS, GHOSTS, INSECTS; *CLAMS, *CRABS, *TURTLES, *SNAILS, ETC.; ARE NOT COUNTED BY X-MAL [UNLESS THE SHELLS ARE REMOVED] #THE GENERAL CLASSIFIER X-UW IS COMMONLY USED FOR THEM. *XNAS *ROOF IS USED WITH X-MAL
-BIUN	1526	GR	*NUMERAL-*CLASSIFIER FOR BREAKS IN THINGS
-GOFET	11431	GR	*NUMERAL-*CLASSIFIER FOR BROKEN OR TORN PARTS OR PIECES
-PILEG	22282	GR	*NUMERAL-*CLASSIFIER FOR BUNDLED OR GROUPED OBJECTS
-YAF	31963	GR	*NUMERAL-*CLASSIFIER FOR BUNDLES OR GROUPS OF TEN (OR EIGHT) REFERRING TO COCONUTS, EGGS, COPRA, ETC.
-PA	21638	GR	*NUMERAL-*CLASSIFIER FOR CHAINED OR STRUNG OBJECTS SUCH AS PALM FRONDS, LEIS, SHELL BELTS
-MATIP	17729	GR	*NUMERAL-*CLASSIFIER FOR CUT PIECES, USUALLY OF VEGETABLES, FRUITS, ETC.
-TIP	20068	GR	*NUMERAL-*CLASSIFIER FOR CUT-UP SEGMENTS (OF BREADFRUIT, TARO, ETC.); #SETIP RE

Figure 6

ENGLISH - WOLEAI FINDERLIST

tube
 pipe, tube:: paip
 tube attached to a stick of dynamite:: raikauq

Tuesday
 Tuesday (lit. second day of work):: Gariuweranel venqaanq

tuq
 pull it loose, pull it off, draw it down, tuq it:: tefinqi
 pull it up, tuq (on) it:: lugosi
 pull it, draw it, pluck it, tuq it, pull it out, take it to pieces, destroy it::
 taiuv
 to pull loose, pull off, draw down, tuq:: tefitef
 to pull, draw, tuq, pluck:: taiuteiu
 to pull, tuq, draw, drag, trail:: luq
 tuq it, pull it, pull and fold it:: tali

tuq-of-war
 tuq-of-war (a kind of game):: tal\$0

tuqqed
 (to be) pulled loose, fall off, drawn down, tuqqed:: tefinqeq

tumble
 to fall (down), be off one's feet, tumble down:: bur\$2

tuna-fish
 king-size tuna-fish:: tanqir
 yellow-fin tuna-fish:: taquv

turkey:: tuuruki

turmeric:: taiq
 to apply turmeric to one's body:: taiq
 turmeric, yellow or orange-colored baby powder:: ranq\$1

turn
 corner, turn:: fat\$1
 make him turn:: qaliuwekiu
 to blow (as of wind), wave, stir, turn:: filefil
 to change, shift, take one's turn, alternate:: kootai
 to face, turn:: taqiu\$1
 to flip, turn over:: volcal
 to spin (many times), rotate, turn:: taquluqul
 to turn (as of a sailing canoe), move around:: sap
 to turn around:: scssor\$2
 to turn around, be turned (of the ends):: sessor
 to turn around, be turned over:: wqiteq\$1
 to turn around, move round:: faan\$1
 to turn one's head:: liuvek
 to turn out, appear, become clear, come into view:: naq\$1
 to turn over, change from primitive ways of life to modern, civilized ones, be
 converted, transferred:: weq
 to turn, be bent, be twisted:: iyaap

Figure 7

beso's	bedesi'1	*oar; *paddle; *propeller;	bab	bebu'1	area/space *above; *top; *surf
bla'i	blil	*house; *household; *family	bad	bedu'1	*asleep
buch	becbi'1	*spouse; *husband or *wife	bar\$1	beru'1	*blanket; *bedding
but	btil	*genitals; *anus; *vagina;	ba'u\$1	bul	*smell; *odor; *scent
cha'lech	chelecheli'1	end of bamboo pole of sail	ba'u\$1	bekebu'1	n.obliq.poss. *smell; *odor; *s
chas\$1	chesi'1	n.obliq.poss. *soot; *ink; *	redu'1\$1	bdelu'1	*head; *leader
ched	chedi'1	low *tide	bedu'1\$2	wobedu'1	n.obliq.poss. *direction; <*fac
chelebla'd	chelebeldi'1	*deception; cheating; *vild	teka'i	bekiu'1	*pottery; clay *pot/*jar; incu
chele'd	cheldi'1	any product of the sea (cr,	bleche's	blechesu'1	new or clean state of somethin
cheltech'a't	cheltecheti'1	wound	blenq'r	blenqru'1	*seal
chese'ches\$1	chesechesi'1	*leprosy; disease with *so	brer	berru'1	*raft (usually made of bamboo)
cheti'1	checheti'1	n.obliq.poss.redup. has com	buk\$1	bkul	*plate; *bowl
chur\$1	cheri'1	*laughter	chab	chebu'1	*ashes; *fireplace; *hearth
dach	decbi'1	*excrement; *shit; *resid-	cha(i)dire'nq	cha(i)direnqu'1	*heart (=internal organ); co
ba'ched	bechede'1	protruding struts on outr	belu'u	belua'1	*country; *village; *place; *te
bachedi'1	bachedile'1	*diarrhea (substance)	bunq\$1	bnqal	*flower; green coconut sheath p
ba'chel	bechele'1	Palauan *money in form of	chad\$1	cheda'1	*alive; living
ba'dek	bedeke'1	*mat (or similar object) ;	chad\$2	chedenqa'1	*liver
ha'eb	bebe'1	*pipe (for plumbing, etc.);	char\$1	chera'1	*price; *cost; *amount of money
ba'il	bile'1	(article of) *clothing; *	cheda'm	deba'1	father (term of address less fo
ba'is	bise'1	action of wandering aroun	cheldechedu'ch\$1	cheldechedecha'1	*conversation; *speech; *
ba'kes\$1	bekese'1	*step (in walking)	chim	chima'1	*hand; *arm; front *paws (of an
ta'kes\$1	blekikle'1	n.r.s.obliq.poss. move (le	chur\$3	chura'1	*tonque
baks	bekse'1	*box (made of any materia	dinq	dinqa'1	*ear
ba'lech	beleche'1	*slingshot; any material	dub	dbal	*bomb; *dynamite; anything *des
banqch	benqche'1	*bite	dach	decha'1	*ability; *skill
bandd\$1	benqde'1	*bounce; *rebound; *suspe	du'i\$2	dia'1	*title (for village chief or fa
banqk\$1	benqke'1	*bank; any storage place	kuteli'nq	kutelnqa'1	*bow of boat; either end of can

Figure 9

F60	ENGLISH - KIRIBATI GLOSSARY COMPILERS 10/10/80 64	ENGLISH - KIRIBATI GLOSSARY COMPILERS 10/10/80 65
	<p>TONGO -- N. A PREPARATION OF XKABUBU] AND MOLLASSES. TOTO/KI -- A SOFT PREPARATION OF XTUAÆ. TUA/E -- DRIED PAN PASTE. TUAÆ -- N. PANDANUS PULP DRIED IN LAYERS ON LEAVES. V.T. XTUAENA]: TO MAKE INTO ... TUAIROA -- V.T. TO GRATE. TO SCRAPE ON GRATER (COOKED PANDANUS). N. XTE KAI N TUAIROA]: GRATER FOR ... TUAMWI -- N. A DISH MADE FROM TARD AND COCONUT. U/XA -- OPEN UP OR UNCOVER AS A NATIVE OVEN AFTER BAK]ING. U/XEBIRI -- STATE OF DOUBT OR PERPLEXITY AS TO THE SUPPLY OF FOOD. U/XINGNA -- EATING OUT THE KERNEL OF A CN WITH THE TEETH. GNAW. U/MUNA -- BAKE IN AN EARTH-OVEN. U/NGIRA -- CHEW AND EAT OR SUCK AS PANFRU OR XBUNIA. UA/NGA -- OVERFEED. UAS -- FRUIT. UA/KANA -- ENGAGE REPEATEDLY IN TAKING FOOD TO A PERSON. UAKANA -- V.T. XUA-KANA]: TO HAVE. OR CARRY PROVISIONS. VICTUALS. UAKANGKANG -- N. A DELICIOUS COCONUT. UB'A/E TE -- SUFFERING WITH DYSPEPSIA OR A SENSE OF FULLNESS AFTER EATING. UKIA/E -- BITE OFF THE XUKIA BEFORE CHEWING THE PANFRU. UNIMU -- FISH SET SIDE TO BE BAKED; BAKED IN (A)P. #1# UN/RA -- EAT OR CHEW PANFRU, XBUNIA MUSK, OR SUGAR CANE. UNG -- CHEW AND EAT OR SUCK PANFRU OR XBUNIA. VA/NIMB/TE -- TO BE VERY HUNGRY. WA/NGA -- TO FEED WELL. WAI N ANG -- (1) IDIOM. XE WAI TE ANG]: THE WIND COMES UP... BLOWS. GUSTS OF WIND. (2) A REFRESHING DRINK. AGREEABLE ... WAIN -- N. WINE. WIKANG/KANG -- HAVING A VORACIOUS APPETITE. WIKANGKANG -- N. GOURMET. GOURMAND. LOVER OF GOOD FOOD. WITA -- N. WHEAT.</p>	<p>ATI TANO] A. SHALLOW, NEEDING A HEAP OF SAND TO COVER. BI/KE -- WHITE SAND, THE SANDY BEACH, SANDY LAND WITHOUT VERDURE. BIKE -- N. BEACH SAND, SAND BANK, SANDY SOIL, THE BEACH, THE SHORE. BU/NGINTA:] -- THE WESTERN HORIZON. BUKINARORA -- N. REEF JUTTING OUT AT END OF LAND. BUKINAEWE -- N. EXTENSION OF REEF BEYOND XBUNINARORA]. BUKININAIWAI -- N. EXTREME EDGE OF REEF. E/KA -- N. ACRE. E/NATINAI -- GO TOWARD OR ONTO THE SHORE, AS FISH WHEN FRIGHTENED. EMSAIA -- N. EMPIRE. I/BENTANO -- CLCD OF EARTH. I/TIMATI -- N. ISTHMUS. IA/NEA -- STRANGER, NATIVE OF ANOTHER COUNTRY. IKUE/TOA -- N. EQUATOR. KA/HARU/ARUA -- VA DEPRESSIONS, OR PITS OR VALLEYS. KABI -- A BAY OR INLET OF SEA OR LAGOON ADVANCING IN POINT OR ROUND INTO THE LAND, A CREEK. KABINO/JUNDU -- A SPACE MORE OR LESS DEEP ON A FLAT. KABIO/NGONGD -- DEEPLY INDENTED AS THE SHORE. KABUAKA -- N. THE DARK, ROCKY BED OF LAGOON, AS OPPOSED TO SANDY BOTTOM. ANT. XKAMAI]. KA NI KAKIKI -- N. A CROSS BAR, A WOODEN TOOL WITH SHARP POINT FOR EXTRACTING LARGE MOLLUSC FROM SHELLS, A SANDBANK EXTENDING UNDER WATER. KAM'A/NOKA -- HAVING A SLIGHT COVE, OR INDENTATION, AS A BEACH. KAM'A/NONO -- A SMALL DEPRESSION OR PIT. KANAMO -- FORMING A POND OR LAKE AT LOW TIDE, OR CREEK. KARA -- N. A LAYER OF HARD SOIL OR SOFT ROCK. (SOMETHING BETWEEN THE TWO). KARANGA -- N. A RIVER. V. TO MAKE FLOW. KARARANG -- FROM XRANGA]: FLOW. N. RIVULET, BROOKLET, SMALL TRENCH, CHANNEL, TRICKLE OF WATER. KE/KE -- A CAPE OF LAND. KEKE -- N. A CAPE, END OF LAND JUTTING OUT TO SEA. KIMARAWA -- N. (1) STALACTITE OF CORAL (GENERALLY RED). (2) A PLANT. M'A/EIET -- EARTHQUAKE, SWAYING. MA/IAKI -- SOUTH. MA/INIKU -- EAST. MA/UNGA -- MOUNTAIN. MAIKU -- N. XMAIKUN TE] ... ON INSIDE OF CURVE OF ISLAND. MANGIEVETE -- OR XWANGIWE TE] A SHARP POINTED CORAL STONE. MANGIENE -- N. A KIND OF CORAL STONE. MANOKU -- (1) N. BAY, GULF, CREEK, ELBOW, COVE. A. XANOKU. XAMANDKU]: CURVED, #L-SHAPED ANGLE, FULL OF COVES. V.A. XAMANDKUA]: TO FORM A ... (2) N. A FISH (XIKAWAIN TE KUAI)]. MARAN -- A. SMOOTH, GLOSSY, POLISHED. MATA N RAWARAWA -- N. AN OPENING, CHANNEL, GAP IN THE REEF. MAUNGA -- N. HILL, MOUNTAIN. A. XAE NAMAUNGA]: VERY MOUNTAINOUS. MBEHE -- N. VENT HOLE OF OCTOPUS. SYN. XNIMAINAI]. N. STALAGMITE (XIBANABA)]. NA -- LOOSE ROCKS OR STONES ON A FLAT. NAMA -- N. LAGOON, LAKE (DIFF. FROM XNAMO)]. NAMO -- N. A SMALL LAGOON, SMALL LAKE, SHALLOW WATER FORMING LAKE AT LOW TIDE IN CERTAIN PARTS OF LAGOON. V.A. XNAM RADI]. FIG. WELL PLACED, NESTLED. NANGANANGA -- N. GROTTO, CAVE, CAVERN, EXCAVATION. NARI -- (1) A LAGOON FISH, SCOMBER, MULLET. XTE NARINARI]: SMALL XNARI. NE/I -- FRESH WATER POOL OR POND. NEI -- (1) N. A POND, POOL, SWAMP, MARSH. NI/KU -- WIDE AS THE RIM OF LAND IN CERTAIN PLACES. NIKURAROA -- N. WIDE PART OF ISLAND, LAND EXTENDING FAR ACROSS THE ISLAND. NU/KAKOTU -- AN UNINHABITED TRACT BETWEEN VILLAGES. NU/KU -- VERY WIDE AS THE LAND. NU/NEABA -- THE MIDDLE OF THE ISLAND RIM FROM OCEAN TO LAGOON SHORE. O/BO -- PLACE NEAR THE BEACH NEVER SHOAL AT LOW WATER. O/NABU -- PORTION OF A CORAL FLAT SLIGHTLY RAISED. O/NATABU/KIBUKI -- UNEVEN, ELEVATED A LITTLE IN PLACES, AS THE SURFACE OF THE GROUND OR OF A PLANK.</p>
-GEO	<p>B/TIM/ANE -- LEDGE OF CORAL ROCK IN THE LAGOON SMALLER THAN A XRAKI. B/TIMA/UNA -- SUNKEN ROCK OR LEDGE WHICH CAN BE SEEN WHEN IT IS CALM. BTIM'A/KORD -- ISLET IN THE REEF. BTIMOTU -- LARGE ROCK IN A LAGOON. 'E/BERA -- N. HEBREW. 'E/RENE -- A GRECIAN. 'I/TA -- N. HEATH. 'ORA/ITAN -- N. HORIZON. A/BA -- LAND. A/BA-NA/KORD -- ISLAND. A/IAKI -- SOUTH. A/INIKU -- OCEAN SIDE OF A CORAL ISLAND. A/NEA -- FOREIGN LAND, LAND NOT ONE'S OWN. A/OBUAKA -- ROUGH, AS LAND, UNEVEN LAND. A/ONTEDRA -- THE FLATS (I XORA] THE FLAT REEF FROM THE BEACH TO THE BREAKERS, A SHOAL). A/ONTIA -- LAND ADJOINING (I XTIA] A BOUNDARY, LIMIT). A/ORAOI -- LEVEL PLACE, NOT RIDGY OR ROUGH, HENCE EASY AS A YOKE, LEVEL. A/RORA -- RIDGE IN THE CORAL REEF ON THE OCEAN SIDE. A:/TIMA:TA -- A SMALL SUNKEN ROCKS. A:/TIMOTU -- A LARGE SUNKEN ROCKS. AN/NA -- TERRA FIRMA. ADATA:/TA -- BEACH ABOVE THE HIGH WATER LINE. ADNA:/BA -- THE EARTH, ALL LANDS. ADNE/INEI -- WET LAND, LAND HAVING WATER BENEATH NEAR THE SURFACE. WET, AS LAND. ATITE/I -- LOFTY CORAL ROCK BENEATH THE SURFACE OF THE WATER. B'A -- A CONTINUOUS HARD CORAL ROCK OR LEDGE. B'A/KATE/KE -- A ROUGH LEDGE OF ROCK OR STRATUM. B'A/NGANING/TANG -- A CAVE. BAKATATI -- N. STEEP SLOPE, CLIFF. A. STEEP, SHEER, CUT STRAIGHT, IN LINE. V.T. XBAKATATA]: TO CUT STRAIGHT, TO CROP STRAIGHT. BAOTI -- N. REEF, SUBMERGED ROCK. BATANO -- N. SAND STONE, CRUMBLING ROCK, SOFT ... SYN. XTE</p>	

Figure 10

a2nBtagiurium0nR nR(aitagiuriu-igR). nCN. nRay back. nIYe biun shiu t. nRMy backbone is broken.
 a2nBtagiurupaa0nR nR(aitagiuriu-paa0nR). nCN. nRa kind of lizard. nIResepaal re mwomwoujo t. nRJapanese people a
 te lizards.
 a2nBtagiuter0nR nR(aitagiuteriuR). nCN. nRa kind of needle fish. nIseuw t. nsemal t. nRa needle fish.
 a2nBtagiyatu0nR nR(aitagiyatuuR). nCVI. ADJ. nR(to be) high, tall, great, lofty. nIYe t. wa we. nRThe canoe is high
 a2nBtagiyetaatn0nR nR(aitagiyataatanR). nCN. nRhigh place, raised place. nIYe matt woal seuw t. nRHe is sitting on a
 raised place.
 a2nBtagomeliiv0nR nR(aitagomeliivuuR). nCN. nRa kind of breadfruit with smooth surface and white flesh. nIYe iyeri
 sefash t. nRHe picked breadfruits from a itagomeliiv nRtree
 a2nBtagowaaaliyel0nR nR(aitagowaaaliyalinR). nCVI. nRto be dizzy, go around in a circle. nIYe sa t. sar we. nRThe c
 hild is dizzy as a result of going around in a circle.
 a2nBtaguluguin0nR nR(aitagulugulu0nR). nCVI. nRto spin (many times), rotate, turn. nIYe t. misil we. nRThe machine i
 s turning.
 a2nBtaguun0nR nR(aitagunuuR). nCVI. nRto turn, return, shift (of wind). nIYe t. tanqi. nRHe turned away from her.
 a2nBtagutog0nR nR(aitagutagunR). nCVN. nRto chase, block. nIYe t. ig. nRThey are chasing fish. nCCF. nBtaguuv.
 a2nBtaguuv0nR nR(aitaguu-anR). nCVT. nRchase it, block it. nIYe t. ig we. nRThey chased the fish. nCCF. nBtagutog.
 a2nBtaguuv0nR nR(aitajuwannR). nCN. nRyellow-fin tuna-fish.
 a2nBtaig0nR nR(aitainR). nCNSP. nRnegativeRaz nRnot. nIYe tai gaang. nRIt is not me. nIYe tai lag. nRHe did not g
 o.
 a2nBtaig0nR nR(aitaiy0nR). nRnyuIrang nBin Paralep dialectRaznn nR. nCN. nRturmeric. nIYe lag tinjar t. we reel
 welewe sin. nRShe went to ask for turmeric from her mother. nB2. nCVI. nRto apply turmeric on one's body. nIYe t.
 sar kave. nRThose children are putting turmeric on their bodies.
 a2nBtaik0nR nR(aitaik0nR). nCN. nRplatform of a house, stones around the house used to keep the gravel from spr
 eading out. nIYe matt wetaif. nRShe is sitting on the platform.
 a2nBtaik0nR nR(aitaik0nR). nCN. nRmountain, hill. nIYe toulap t. woal Hawaii. nRThere are many mountains in Havi
 i.
 a2nBtaikail0nR nR(aitai-kailanR). nCVI. ADJ. nR(to be) weak, unhealthy. nIYe t. nRweak person. nCCF. nRkail.
 a2nBtaik0nR nR(aitaik0nR). nCJAP. nCN. nRdrum. nIYe sa gakiungiu t. we. nRThey have sounded the drum.
 a2nBtaikusang0nR nR(aitaikusanganR). nCJAP. nCN. nRcarpenter.
 a2nBtaik0nR nR(aitaik0nR). nCVT. nRsharpen it, make it sharper. nIYe t. saat we. nRHe sharpened the knife. nCCF.
 nRtaitei.
 a2nBtaing0nR nR(aitainganR). nCVI. nRto menstruate, be in one's menstrual period. nIYe sa t. Shoabut ve. nRThe wom
 an is in her menstrual period.
 a2nBtaisob0nR nR(aitaisob0nR). nCJAP. nRnysententialRaz nRDon't worry! No trouble!
 a2nBtaitei0nR nR(aitaitai0nR). nCVN. VI. nRto sharpen, make sharper. nIYe t. saat. nRHe is sharpening a knife. nIY

Fig. 11

AVUD0	AVUDU	AVUGA	AVUGE	AVUGI	AVUGO	AVUGU	AVUGWA	AVUG'A	AVUG'E
AVUG'I	AVUG'G	AVUG'U	AVUHA	AVUHE	AVUHI	AVUHO	AVUHU	AVUKA	AVUKE
AVUKI	AVUKU	AVUKU	AVUKWA	AVULA	AVULE	AVUET	AVULC	AVULU	AVUMA
AVUME	AVUMI	AVUJU	AVUMU	AVUNA	AVUNE	AVUNI	AVUNO	AVUNC	AVUPA
AVUPI	AVUPI	AVUPU	AVUPU	AVURA	AVULC	AVURI	AVURG	AVURU	AVUSL
AVUSI	AVUTA	AVUTU	AVUTU	AVUVA	AVUVE	AVUVI	AVUVU	AVUVL	EBABA
EBABE	EBABI	EBABD	EBABU	EBADA	EBADE	EBADI	EBADD	EBADU	EBAGA
EBAGE	EBAGI	EBAGU	EBAGU	EPAGWA	EBAG'A	EBAG'L	EBAG'U	EBAG'U	EBAG'U
EBAHA	EBAME	EBAMI	EBAMU	EBAMU	EBAKA	EBAKE	EBAKI	EBAKL	EBAKU
EBAKWA	EBALA	EBAL	EBALI	EBALO	EBALU	EBAMA	EBAM'E	EBAMI	EBAM
EBAMU	EBANA	EBANE	EBANI	EBANO	EBANU	EBAPA	EBAPE	EBAPI	EBAPU
EBAPU	EBARA	EBARE	EBARI	EBARO	EBARU	EBASE	EBAGI	EBATA	EBATO
EBATU	EBAVA	EBAVE	EBAVI	EBAVU	EBAVU	EBEBA	EBEUE	EBECI	EBEBU
EBEGU	EBEDA	EBEDE	EBEDU	EBEDU	EBEDU	EBEGA	EBEGE	EBEGI	EBEGO

Fig. 12

<p>191 λ'u'e tu' xe'e he s/pu'λ'm-s tu + /ci'txw and.so from NEARBY DIR NOB/exit-3.PSV from(v.p.) E.P /house and then he went out of the house</p>	<p>MM40 † /xwuy' /xwesi't-s-t-n us E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV that thing which was going to take him</p>
<p>23 /p'e'n't-s-t-e xe'e w + e c'i'e' vs. /Return-CAU-TR-1SP NEARBY toward E.P DIR EMPH/INTRO CJV Put that back here where it belongs!"</p>	<p>MM38 † /xwuy' /xwesi't-s-t-n us. E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV that thing which was going to take him.</p>
<p>16 /xwu'y' xe'e /Xək-p-s-t-e'ne /FUT NEARBY /know-1NC-CAU-TR-1S.SBJ "I'm going to find out</p>	<p>MM31 † /xwuy' /xwesi't-s-t-n us E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV what he was going to travel in</p>
<p>/xe'-t-/tai'xw</p>	<p>xwas=xwesi't</p>
<p>32 u † /xe'-t-/tai'xw toward E.P /above-LIG-/world to go up to heaven--</p>	<p>MM2 λ'u'e s/xwu'y'-s xwas/xwesi't and.so HCR/FUT-3.PSV AUG+travel who was going to take a trip</p>
<p>/xe'-t-/tai'xw.</p>	<p>/xwu'y'</p>
<p>33 u † /xe'-t-/tai'xw. toward E.P /up-LIG-/world up to heaven.</p>	<p>MM72 'u'u, /xwu'y' kn /kwne'-n tək /pe'ye' oh /FUT I /take-HDL DESCR /one "Oh, I'm going to take one</p>
<p>34 u † /xe'-t-/tai'xw. toward E.P /above-LIG-/world to heaven.</p>	<p>MM16 /xwu'y' xe'e /Xək-p-s-t-e'ne /FUT NEARBY /know-1NC-CAU-TR-1S.SBJ "I'm going to find out</p>
<p>/xe'-t-/tai'xw.</p>	<p>/xwuy'</p>
<p>41 u † /xe'-t-/tai'xw. toward E.P /above-LIG-/world up to heaven.</p>	<p>MM74 /xwuy' kn /kwne'-n tu' xe'e tək / /FUT I /take-HDL from NEARBY OBL UMR /one I'm going to take one of these."</p>
<p>/xi'ep'e'y'us</p>	<p>MM33 † /xwuy' n/xwesi'(t)-ta-s E.P /FUT LOCL/travel-1NS-3.PSV his intended means of traveling</p>
<p>89 /te'kn + /xi'ep'e'y'us /all E.P /shelf It was all of the shelves</p>	<p>MM8 tək /c'i'y us /xwuy' s/cu-tə'n-s. DESCR /certain-way CJV /FUT WOH/do-1NS-3.PSV about how to do things.</p>
<p>87 † /xi'ep'e'y'us u ci'e' E.P /shelf toward THERE --the shelves there--</p>	<p>MM38 † /xwuy' /xwesi't-s-t-n us. E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV that thing which was going to take him.</p>
<p>/xi'ep'e'y'us.</p>	<p>MM31 † /xwuy' /xwesi't-s-t-n us E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV what he was going to travel in</p>
<p>78 tu † /c'y'e' tu + /xi'ep'e'y'us. from(v.p.) E.P /basket from(v.p.) E.P /shelf from [where] the baskets [were]-- from the shelves.</p>	<p>MM40 † /xwuy' /xwesi't-s-t-n us E.P /FUT /travel-CAU-TR-(3.OBJ)IDF CJV that thing which was going to take him</p>
<p>/xw'-i'...ot</p>	
<p>56 λ'u'e s/vi'k-c /xw'-i'...ot † c'i'e'/c'y'e'. and.so HCR/see-TR.3.SBJ /many-PHY E.P AUG-/basket and he saw many, many baskets.</p>	

Figure 13

Hsu, R. "The role of computers in dictionary-making at the University of Hawaii". In Pankey, A. and Carrington, L. editors, *Austroasiatic Linguistics at the 15th Pacific Science Congress*. C-88-313-328. Pacific Linguistics, The Australian National University, 1985. DOI:10.15144/PL-C88-313
©1985 Pacific Linguistics and/or the author(s). Online edition licensed 2015 CC BY-SA 4.0, with permission of PL. A sealang.net/CRCL initiative.

THE POLITICS OF DICTIONARY MAKING ON TANNA (VANUATU)

Lamont Lindstrom

Sturtevant has defined a culture as "the sum of a given society's folk classifications" (1964:100). This interpretation of culture – although it can be taken as a gross simplification – stresses the centrality of socially constructed definitions of reality. Dictionaries, in literate societies, are folk attempts to standardise a society's classifications and definitions. They are part of the apparatus by which cultural knowledge is codified and transmitted. Codification systematises cultural definitions and their linguistic labels. Transmission ensures that the systematised cultural code extends throughout a society and across time.

Codification and transmission of standardised cultural definitions are not apolitical processes. Instead, they forward the interests of some people and groups and challenge those of others. At the broadest level, political competition involves definitions of reality. Competing groups advance variant interpretations of the world. Concepts (e.g., of natural and unnatural, masculine and feminine, wisdom and stupidity, goodness and evil) must be continuously validated (and sometimes revised) in social interaction and argument. Those individuals and groups commanding positions of political and economic power within a society also control the cultural definitions of that society, and their codification as transmitted by dictionaries.

Powerful groups validate and maintain their command of social reality by codifying and transmitting this in dictionary form. The appearance for the first time of authoritative English dictionaries in the 18th century (Wells 1973) correlated with increased political muscle of the British middle class. The programmatic statements of early dictionary makers and their supporters castigated the speech of both the vulgar poor and "people of fashion" (Wells 1973:46). The more recent publication of Webster's *Third international dictionary* – which for the first time listed and defined "ain't" and a number of other rude American words – occasioned a long debate about the authoritative versus descriptive functions of dictionaries (Sledd 1962). Those who protested the vulgarising of dictionary language accurately perceived the political competence of dictionaries which protect dominant group interests by making a particular speech style and system of folk classification the standard.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 329-341. *Pacific Linguistics*, C-88, 1985.

© Lamont Lindstrom

The question comes down to the degree of shared culture (including language) within a society. If a culture is entirely shared, no disagreement or conflicting interpretations of word meaning or of word pronunciation could exist. A dictionary would be completely descriptive and this description would have no political significance. Much of culture, however, is not shared. Groups and individuals within a society possess different sets of definitional constructs and work with variant grammatical rules which generate a number of different speech styles. Most speakers, of course, agree on at least the primary codified meanings and indicated pronunciations of many of the words found in an English dictionary. One still need ask, however, after determining that culture is shared to some degree, how these particular codifications become and remain standardised.

Dictionaries transmit an interpretation of reality. Even if they succeed in being partially descriptive of shared and variant cultural meanings and linguistic form, they remain authoritative political statements. A dictionary is authoritative not only in the sense that it instructs its readers in the correct manner of defining, pronouncing and spelling, but also because the particular definitions and speech styles it codifies and transmits become a standard removed from ongoing speech interaction. By codifying a standard code and by storing and circulating this in literate form, dictionaries objectify language. Thus objectified, dictionary-disciplined language achieves greater autonomy than language which exists only in memory.

If all dictionaries demand "making and controlling translations" (Voegelin, quoted in Robinson 1969:10), bilingual dictionaries involve further, cross-cultural considerations of control. Writing bilingual dictionaries is a small part of Western appropriation of the world. Linguistics, like anthropology, fixes in print a cultural system in such a way that it becomes a knowable object more accessible to manipulation by those both within and without the speech community. Malinowski, who instituted early anthropological and linguistic fieldwork in the Pacific, scribbled in his diary as he sailed north to the Trobriand islands:

I hear the word "Kiriwina" ... I get ready; little grey,
pinkish huts ... It is I who will describe them or
create them (1967:140).

His claim, grandiose and egomaniacal, nevertheless applies also to the composition of dictionaries. Dictionaries objectify sounds into orthography, utterances into morphemes and inference into denotation. Dictionary codification is literary cryogenics. In addition to capturing only a particular moment in communicative flux, a dictionary flash-freezes a language into a configuration which is only one of a number of possible abstractions of its present state. Each of these alternatively possible dictionary codifications presents its own attendant political implications.

This paper discusses three codificatory puzzles which arose in the compilation of a dictionary of the N̄nin̄ife (Kwamera)¹ language of Tanna in the southern part of Vanuatu (Lindstrom forthcoming). These puzzles consist of island words which are more than arbitrary acoustic symbols of material and immaterial ideas. They also indicate something about the speaker and speech context. This secondary, political utility often dominates the primary referential function of a word (cf. Salisbury 1962; Strathern 1975; Sankoff 1976, 1977).

About 17,000 people live on Tanna and speak five closely related Austronesian languages (see Lynch 1978; Tryon 1976). Two thousand people along the south and east coasts of the island speak N̄ninife (described in missionary sources as Kwamera). Presbyterian missionaries, during the late 19th and early 20th centuries, invented several orthographies of three of the island's languages acting according to the usual protestant dictum of Bible translation. They produced a N̄ninife New Testament as well as a number of hymnals, elementary primers and other material used in mission schools (see Watt 1880, 1890, 1919, for example). A generation of men, now in its late 50s and 60s, learned to read (more than write) their language.

In the 1960s, the British and French colonial governments took control of and expanded the mission school systems. Political concerns in the main motivated this educational expansion. Government schools purposely neglected indigenous languages as well as Bislama, the Pidgin English lingua franca of the archipelago, to ensure student literacy in one or the other of the colonial languages. Few young Tannese can read their own languages, although some have a passing acquaintance with English or French.

Although the recently independent Vanuatu government supported a language conference in 1981 which made recommendations concerning the future role of the nation's 105 indigenous languages in education, law, and the mass media, it has yet to undertake much of a program to ensure their national significance or utility. The conference did recommend, however, the production of dictionaries partially as linguistic salvage (of those languages "on the verge of being lost because of declining population") and partially to transform (literalise) local languages into objects of utility within national institutional contexts (Pacific Churches Research Centre 1981:17). This dictionary objectification of local languages is an initial requirement for subsequent national appropriation and manipulation.

Dictionaries make sense by codifying word meaning and word form. Attempts to codify local languages, however, encounter a number of practical problems with serious political implications. Some difficulties relate to the fact that word meanings are socially unshared. Other difficulties relate to variant word form. This paper discusses the problematic codification of three sorts of politically significant words. Some words are meaningful because they have no meaning. These function, partially, to signify the importance of a communication. Others are words the articulatory rights to which individual speakers inherit and control. These mark personal distinctiveness and identity. Finally, a third type of words consists of sets of cognates which are associated with particular residential groups. These words symbolise speakers' local affiliations and signify the existence of political boundaries.

Dictionary codification flounders in the first instance in that although speakers use a word they do not share its meaning. It flounders in the latter two instances in that although speakers share meaning they are unable or unwilling to pronounce the word. Because of the significance of these variations, the choice by a dictionary maker to resolve codificatory incertitude in one way or another may have local political impact if his dictionary becomes known and used.

WORDS WITHOUT MEANING

People sometimes use words the meaning of which they claim not to understand. Malinowski, encountering similarly senseless words in Trobriand Island garden spells, described the problem as "the meaning of meaningless words" (1935:213). Nonsensical words, on Tanna, occur principally in song (cf. Fortune 1963:257-258; Lewis 1980:59). People discern songs to be ancestral messages. These may be inherited from forebears or have more immediate origins if some songsmith is ancestrally inspired as he dreams. Gray, a 19th century Presbyterian missionary on Tanna, noted:

a native, we know, readily uses the preformatives of his own dialect with the stem root words of another dialect. I have found these corruptions and foreign words in all native songs I have examined (1894:43, see also Codrington 1891:334-336).

"Meaningless word", of course, is an oxymoron. A nonsense word has meaning even if this is inferential rather than referential. Malinowski suggested that meaningless words function to mark the extraordinariness and magical status of an utterance (1935:224). The words of Tannese songs, partially or completely senseless, share this utility. Songs are the chief form of ritual speech at traditional ceremonial occasions. Supporters of the principals involved in the day's exchange of goods gather to dance and sing throughout the night. Singers are ignorant of the sense of many of the traditional songs in their repertoire. Meaningless libretti also characterise the songs which people sing during the ceremonial events of modern ideological organisations. These include the various Christian sects and the John Frum Movement (a successful political organisation cum cargo cult). Christians, for example, are content to yodel English or French hymns, singing words with no denotation for most of the hymnists.

* * * * * * *	
lou, Tomi Timi Kaupoī Tīna Okei okei Wel tumaruma Iso soera tieni tenama.	<i>I, Tommy, Jimmy Cowboy Tanna Okay, Okay (senseless language until song's end).</i> John Frum Hymn
* * * * * * *	

An equation of semantic opacity, remoteness, and antiquity informs folk etymology. People, to account for their choral lexical ignorance, suggest that nonsense words either are of foreign origin ("Tahiti" and "Tonga" are suspected venues) or are the speech of the ancestors. In some cases a word may be both these things; linguistic consultants sometimes identify a word which exists as a common form in a neighbouring dialect as ancestral, and therefore spookily senseless.

Even though Malinowski claimed that nonsense words are meaningful "in that they play a part" (1935:247), he was also very concerned to pin down any denotations he could. He relied sometimes on flimsy morphological evidence but more often on his key informant in these matters, Bagido'u:

In some formulae we are able to translate the words clearly and satisfactorily after our magically illumed commentator has given us their esoteric meaning (1935:219).

Malinowski's anthropological efforts, however, to elucidate and codify these meaningless lexical riddles run counter to politically functional ambiguity in Melanesian societies. A word meaning known by a single person (i.e., "wise informant") is not a social fact until this meaning is communicated to another. The transformation of personal interpretations into socially shared meanings is one of the bases of power in the area. On Tanna, this exegesis of esoterica is the main avenue to prestige (Lindstrom 1984). The existence of political competition on the island generates much more disagreement than agreement in semantic interpretation. Malinowski, had he found a second wise informant, would probably have discovered likewise divergent explications (see Malinowski 1935:232; Lewis 1980:67-71).

Songsmiths, on Tanna, continue the production of nonsensical songs in order to sustain an interpretive role. Nikiau, for example, a John Frum leader of the 1940s, instructed young men and women in the meaningless words of a set of new cult songs. These represented, he claimed, John Frum's language. He instantly became a religious pundit and an individual of some prominence in as much as people were willing to sing the songs according to his interpretations. Meaningless words provide material for politically motivated exegesis. A particular semantic interpretation, of course, may or may not establish much exchange value. A semantic savant's political success within the local information market is measured by the degree to which his interpretations are accepted by the public (and, sometimes, by his ability to convince — or take in — visiting ethnographers such as Malinowski). Dictionary codification of one interpretive version of these words obviously would lend support to one leader vis-a-vis his semantic competitors.

A leader, or big-man, in this sense is an interpreter. He, too, is a dictionary maker. His advantage is that his interpretive codifications are stored in memory rather than in print. The 'meaning' of this sort of word is socially constructed to a degree far beyond the imagination of any phenomenologist. These meanings have no guarantee of permanency, depending as they do on political exigency, and they thus violate the temporal semantic expectations which make dictionary making possible. Meaningless words, which signify the specialness of a communication or permit definitional fancy, must be glossed as political supersense rather than nonsense. This sense, however, will probably decay before a dictionary does.

VERBAL ASSETS

There are words which everyone speaks and sings but only certain people agree to understand (as above). There are others which everyone understands but no one speaks. The problem with this second category of word is not the codification of meaning; it is a problem of word control. These lexemes are personal property inherited from one's ancestors. Dictionary appropriation of this sort of word becomes a form of symbolic thievery, etyma-larceny, as it were. Fortune, collecting on the sly Dobuan spells which contain various secret names for supernatural actors, animals, things, etc., noted that had he used such names publicly, he

would have aroused such resentment in my teacher of magic that my learning of magic would have been over. I would have been giving names of power, giving power itself, to those who had no birth-right to such power, but who had to fee the special practitioners and possessors of such power to exercise it on their behalf (1963:114).

Personal names are one possible set of verbal assets. On Tanna, as elsewhere in Melanesia, many personal names (which also label plants and animals) belong to particular lineages (or "name-sets", see Lindstrom 1985) and are recycled through the generations. Other islanders, however, have rights of pronunciation of these lineage nomenclatural assets and can use them to refer to the so-named people as well as to their natural object namesakes. The proprietary assumptions linking a person and his name do not entail a speech taboo which prevents the articulation of the name by others (as occurs elsewhere in the Pacific, see Fortune 1932:62-68, for example).

- * * * * *
- K+MTI N
1. *Kind of taro.* 2. *Personal name.*
- K+RA N
1. *Kind of tree.* 2. *Ladder.* 3. *Personal name.*
- PAUPAUK N
1. *Butterfly.* 2. *Personal name.*
- * * * * *

Other verbal assets, however, do entail enunciatory taboos, or at least a wariness on the part of those speakers with no rights to the word. These, especially, are words which label or describe various magical paraphernalia. Most men have inherited magical objects (e.g., sets of powerful stones), along with knowledge of necessary bark and leaf accoutrements and the right to legitimate magical practice. This distinctive knowledge is part of the constitution of every man's individuality. Its transmission is highly restricted in order to maintain its secrecy. On Tanna, there is thus an 'organic' distribution of magical knowledge in which every person controls a small part of the whole. As event dictates, various individual practitioners are called to the fore in order to regulate the weather, diagnose and cure disease, ensure the fertility of the season's crops, etc.

People are conspicuously careful not to violate the barriers of information transmission which would threaten the current distribution of restricted knowledge. Part of this prudence extends to an unwillingness to pronounce in public words associated with one or another of the magical technologies. These techniques frequently involve very similar materials distinguished only nomenclaturally. A magically treated length of wild cane (ordinarily nig) can take a different name depending on which person's magic so treated it. People without rights to operate a magical technique publicly claim ignorance of all that it entails. They reveal only in private their illicit knowledge of associated names and words.

- * * * * *
- NUKWEI NARI N
Sorcery, or magical stone.
- KWATIUTIU N
Magically treated length of wild cane (Miscanthus sp.)
- PWIP N
Magically treated length of wild cane.
- * * * * *

Malinowski, collecting his spells, encountered a similar distribution of verbal assets in the Trobriands. People informed him:

"This is Bagido'u's magic — we cannot speak about it."
It is bad form to trespass on the magician's exclusive field of knowledge (1935:225).

Malinowski went to Bagido'u, learned from him, and subsequently revealed his knowledge in print. Makers of dictionaries need to discern whether speakers make a distinction between oral and written revelation of verbal assets. If none exists, the lexicographer must consider seriously the consequences of potential semantic trespass — a dictionary redistribution of linguistic private property.

There are other words which everyone understands but is wary of speaking, although for different reasons. Here, words are taboo not because they are associated with a body of personally managed secret knowledge but because they serve to mark particular categories of social relationships (cf. Goodenough and Sugita 1980:1-11). A speaker's avoidance of certain words when communicating with an interlocutor marks the social identities involved in the interaction and makes a comment on the current state of the relationship.

Brothers and sisters (real and classificatory), in particular, avoid discussion of topics running the gamut from copulation, through pregnancy to parturition. They are also careful not to use any of a set of marked words which denote sexual body parts and their functions. Men, particularly young men who call each other by a reflexive kin term *-ieri* (*actual/potential brother-in-law*), on the other hand, regularly bandy these terms as part of expected verbal abuse. Violation of either expectation of linguistic immoderation or punctilio signifies some derangement in the social relationship.

*	*	*	*	*	*	*	*
KANARI	N						
<i>Vagina.</i>							
KWANIHI-	N						
<i>Penis.</i>							
KWANARE-	N						
<i>Testicle.</i>							
-EHI	V						
<i>Copulate.</i>		-EHI	IKOU,	<i>copulate from the rear.</i>			
*	*	*	*	*	*	*	*

The conversational exchange of marked words of this sort is also characteristic of various social situations. A major setting for jocular obscenity, for example, is an informal football game during which youthful players comment both on the play of the game and on the qualities of fellow players. Men seemed to experience a certain illicit diversion in teaching me the set of marked vocables and explaining the niceties of their usage. This, however, only occurred within uneasily stimulated all-male groups. Linguistic consultants, nervous at my writing all this down, specifically stated that such words do not belong in a dictionary. (They agree, in this, with Webster.) Dictionaries, unlike football games, ought to contain only polite language.

A comprehensive dictionary could offend people's sensibilities in that it threatens the expected distribution of linguistic markers of social relationships. If people of the wrong kin type in future happen together to peruse the dictionary and encounter a marked term, social tumult akin to an infamous local showing of David Attenborough's film on the John Frum Movement is not inconceivable. Attenborough had photographed men drinking kava — an activity at least ideologically never seen by women. When his film made its way back to the island to play to a mixed sex audience in a school room served by an electric generator, men leapt to their feet in dismay and set about stuffing their wives and daughters under the nearest chairs or hustling them out of the room. Like an ethnographic film, a dictionary — because it is literary — at least partially removes a language from the control of its speakers.

LINGUISTIC CHAUVINISM

A third type of politically significant words consists of limited sets of microdialectal cognates. These words signify speakers' wider affiliations and mark group boundaries (Grace 1981:153-161; Lindstrom 1983). People conversationally recognise the distribution of these cognates to situate speakers within neighbourhoods. (There is also a much larger set of cognate lexemes in free or microdialectal variation throughout the area which people ignore as inferentially useful.) Whereas with senseless words, a dictionary fails in the codification of meaning, here the difficulty is in codification of phonetic form. Although every Nininife speaker knows all significant variants (and can locate these geographically), each uses the set associated with his particular village (cf. Salisbury 1962; Gumperz 1978:394). To do otherwise would signify displacement from his local group.

	* * * * * * *
-ATA	V <i>See, look (also -ATONI, Port Resolution).</i>
-+K+NEK+N	A <i>Strong, rigid (also -+KM+K+N, mountain area)</i>
-KAF+K	G <i>First person singular possessive marker for certain semi-alienable nouns (also KOK-, Imaki area).</i>
REK+M	I <i>No (also REKAKU, Port Resolution; N+K+M, mountain area).</i>
-VEHE	V <i>Come, move towards (also -AFE, Port Resolution).</i>
	* * * * * * *

Speakers of all microdialects claim their particular variant as the 'stump' of language — the origin of all other (distorted) island languages and the proper manner of speaking. They accuse others of misspeaking or twisting real language. Islanders, except in multilingual or joking contexts, avoid producing available variants from other areas (although they understand these) not only because of the symbolic displacement of identity thus generated, but because they consider such variants as outlandish, less prestigious, if not also incorrect. Similar linguistic chauvinism also characterises people's estimations of the island's other languages.

To concentrate dictionary effort on one N̄nin̄ife microdialect would confirm one local group in its prejudices and offend all others. To include all microdialectal variation would please nobody. The exigencies of field-work and personal knowledge, nevertheless, dictate an intermediate course (cf. Harrell 1967:56-57). This involves concentration on one microdialect supplemented with available information from the others (which will, perhaps, both displease and offend).

LANGUAGE OUT OF CONTROL

Writers of bilingual dictionaries must select their audience in order to determine how best to structure the information they compile (Haas 1967). This becomes problematic when recording unwritten languages. In whose society will the controlled linguistic object become a meaningful artifact? If a dictionary has meaning only within one of the societies of the bilinguistic conjunction (the English-speaking), codificatory difficulties which stem from the political utility of language – constantly revised in an arena where political interest partially dictates semantic and phonetic structure – are unimportant. One society's political tool becomes the other's curious artifact and this is acceptable whether or not it contains non-denotative words, individual verbal assets, taboo words, or verbal markers of local group affiliation. If a dictionary, however, becomes a meaningful artifact in both societies, the translations it makes and controls become one of many possible political statements. This dictionary statement differs from the rest, however, in its literate form and permanency. Language, thus codified, escapes the usual controls of individual interests and memory.

Tannese cultural definitions and speech patterns are currently codified only in memory and transmitted by speech. A dictionary constitutes a channel for knowledge codification and transmission which is more powerful than speech, more permanent than memory. Dictionaries, because of this, partially remove a language from the control of its speakers. What was constantly negotiated in political interaction is now frozen in literate form.

Goody and Watt distinguish controlled (or literary) language from non-literate. As characteristic of the second, they argue:

There can be no reference to 'dictionary definitions', nor can words accumulate the successive layers of historically validated meanings which they acquire in a literate culture. Instead the meaning of each word is ratified in a succession of concrete situations, accompanied by vocal inflexions and physical gestures, all of which combine to particularise both its specific denotation and its accepted connotative usages (1968:29).

Dictionaries, because they transcend the control of individual memories and interests, make apparent inconsistencies in language over time and across a society. They make apparent the fact that culture is not totally shared and that language is variable. N̄nin̄ife has changed enough since the publication of a 19th century translation of the New Testament that the men able to read the remaining specimens of this book recognise and comment on the variation. This diachronic variation, however, fits neatly with the idea that ancestral language as spoken either by one's grandparents or by ancestors who appear in dreams should be different from everyday speech.

Dictionary codification and revelation of contemporary linguistic variation (i.e., culture which is either unshared or differentially valued), on the other hand, is more disturbing. A dictionary reveals some of the infrastructure of power and inequality on the island. Moreover, dictionary control of meaningless words, verbal assets, and variant cognates is an objectification of only one of a number of competing political statements. By taking the making and controlling of translations out of everyday interaction, a dictionary as a new artifact in Tannese society could support the definitional claims, political interests, and linguistic expectations of some groups and individuals over others. The dictionary regulation of language may have political consequence as well, in the case of bilingual dictionaries, between societies.

In literate societies, speakers are no longer the sole judge of the meanings and the proper forms of words; nor are they any longer solely responsible for codification and transmission of their language. Dictionaries, instead, define a standard and, therefore, help to reproduce as well as merely describe shared culture. Two hundred years ago, when dictionaries were created to be authoritarian statements of one particular interpretation of linguistic and cultural reality, speakers of English lost partial control of their language.

Standards of meaning and of pronunciation also exist in non-literate (or functionally non-literate) societies such as Tanna. These standards, however, are not predominant in that all speakers negotiate and transmit them daily in public conversation and store them only in memory. There is no determining, written authority. Ruling structures of political inequality, of course, affect the outcome of these processes of conversational negotiation which create and validate shared meanings. Literate dictionary storage and circulation of lexical meanings and forms, however, offers a new mechanism of language control of a different, more durable order. This authoritative competence is given in the name; dictionary, dictum, and dictate, of course, are etymological kin.

This is not to say that a system of defined meanings is immune from challenge because it is written. Speakers, in the end, are capable of regaining a measure of linguistic control by recognising that dictionaries, as authoritative standards, are also political statements. This has already occurred, on Tanna, with ethnographic codifications of non-linguistic aspects of culture (cf. France 1969). In the early 1950s, the anthropologist Jean Guiart attempted to record the names of men possessing rights to two traditional 'chiefly' statuses in every local group. Although ideologically inherited through patrilineal links, men actually appropriate these statuses by astute political manipulations including the revision of the unwritten past. When men peruse this catalog of chiefs today, they are confounded by what they see as a pack of lies. Guiart (1956) remembers in print what they find convenient to forget. They do not, naturally, cease to forget. Instead, Guiart becomes the gullible victim of past deceptions.

A dictionary, as representative of certain interests over others, perhaps expects no better future than codified ethnography. Political circumstance will determine the future standing of its controlled word meanings and phonetic forms. In one event, a dictionary will be a valuable treasury of ancestral speech; in the other, a fraudulent counterfeit of real language.

NOTES

I would like to thank Fulbright-Hays, the English-Speaking Union of the United States, the Departments of Anthropology at the University of California, Berkeley and at the Research School of Pacific Studies, Australian National University, the University of Tulsa, and all friends on Tanna for the assistance I received and the welcome I experienced during three research trips to Vanuatu. I also thank A.K. Pawley for helpful editorial criticism.

1. The symbol [ɨ] represents a mid central vowel; [v] a voiced high central glide; and [ŋ] a velar nasal stop (see Lynch 1978).

BIBLIOGRAPHY

CODRINGTON, R.H.

- 1891 *The Melanesians: studies in their anthropology and folklore.* Oxford: Clarendon Press.

FORTUNE, R.F.

- 1932 *Sorcerers of Dobu.* New York: E.P. Dutton.

FRANCE, Peter

- 1969 *The charter of the land: custom and colonization in Fiji.* Melbourne: Oxford University Press.

GOODENOUGH, Ward H. and Hiroshi SUGITA

- 1980 *Trukese-English dictionary.* Philadelphia: American Philosophical Society.

GOODY, Jack and Ian WATT

- 1968 The consequences of literacy. In Jack Goody, ed. *Literacy in traditional societies*, 27-68. Cambridge: Cambridge University Press.

GRACE, George W.

- 1981 *An essay on language.* Columbia, SC: Hornbeam Press.

GRAY, W.

- 1894 A song of Aniwa. *Journal of the Polynesian Society* 3:41-45.

GUIART, Jean

- 1956 *Un siècle et demi de contacts culturels à Tanna.* Paris: Musée de l'Homme.

GUMPERZ, John J.

- 1978 Dialect and conversational inference in urban communication.
Language in Society 7:393-409.

HAAS, M.

- 1967 What belongs in the bilingual dictionary? In Householder and Saporta, eds 1967:45-50.

HARRELL, Richard S.

- 1967 Some notes on bilingual lexicography. In Householder and Saporta, eds 1967:51-62.

HOUSEHOLDER, Fred W. and Sol SAPORTA, eds

- 1967 *Problems in lexicography*. Bloomington: Indiana University.

LEWIS, Gilbert

- 1980 *Day of shining red*. Cambridge: Cambridge University Press.

LINDSTROM, Lamont

- 1983 Say what? Language and political boundaries on Tanna (Vanuatu).
Anthropological Linguistics 25:387-403.

- 1984 Doctor, lawyer, wise man, priest: big-men and knowledge in Melanesia.
Man (n.s.) 19:291-309.

- 1985 Personal names and social reproduction on Tanna, Vanuatu. *Journal of the Polynesian Society* 94:27-45.

forth-coming *Kwamera dictionary/Nikukua sai nagkiariien Nininife*. PL, B-95.

LYNCH, John

- 1978a *A grammar of Lenakel*. PL, B-55.

MALINOWSKI, Bronislaw

- 1935 *Coral gardens and their magic* (Volume 2). New York: American Book Co.

- 1967 *A diary in the strict sense of the term*. New York: Harcourt, Brace and World.

PACIFIC CHURCHES RESEARCH CENTRE

- 1981 *Report of the Vanuatu Language Conference*. Port Vila, Vanuatu.

ROBINSON, D.F.

- 1969 *Manual for bilingual dictionaries*. Santa Ana, CA: Summer Institute of Linguistics.

SALISBURY, R.F.

- 1962 Notes on bilingualism and linguistic change in New Guinea.
Anthropological Linguistics 4:1-13.

SANKOFF, G.

- 1976 Political power and linguistic inequality in Papua New Guinea.
In W.M. O'Barr and J.F. O'Barr, eds *Language and politics*, 283-310.
The Hague: Mouton.

SANKOFF, G.

- 1977 Multilingualism in Papua New Guinea. In S.A. Wurm, ed. *New Guinea area languages and language study* (volume 3), 265-307. PL, C-40.

SLEDD, J.

- 1962 *Dictionaries and that dictionary*. Chicago: Scott, Foreman.

STRATHERN, Andrew

- 1975 Veiled speech in Mt. Hagen. In M. Bloch, ed. *Political language and oratory in traditional society*, 185-203. New York: Academic Press.

STURTEVANT, W.C.

- 1964 Studies in ethnoscience. *American Anthropologist* 66/3(2):99-131.

TRYON, D.T.

- 1976 *New Hebrides languages: an internal classification*. PL, C-50.

WATT, William

- 1880 *Primer*. Tanna: Mission Press.
1980 *Nakukua savei iarumanu*. Glasgow: Robert Maclehose.
1919 *Naresien em nupume ya nafwakien ya nagkirien Kwamera*. Melbourne: Brown, Prior.

WELLS, R.A.

- 1973 *Dictionaries and the authoritarian tradition*. The Hague: Mouton.

RAPID LEXICAL CHANGE AND THE PROBLEM OF WHAT TO INCLUDE IN A NEW TAHITIAN DICTIONARY

Jack H. Ward

The task of this paper is to address the desirability of including non-contemporary words in a proposed dictionary of Tahitian and the exploration of a certain approach to that problem as well as the resolution of some lexicographical difficulties which arise with the execution of the procedure.

BACKGROUND

Tahitian has been described as being lexically aberrant in comparison with other, closely related Polynesian languages (Henry 1928; LeMaître 1973; Levy 1970; White 1967). This characterisation has been attributed to rapid innovation in vocabulary which in turn involved (1) the loss of aboriginal cultural practices and social institutions in connection with a sudden, intense, and relatively long interaction with Europeans, (2) the eager acquisition of European technology as well as cultural and social traditions (Vernier 1948) both factors being characterisable as the result of external influence, and (3) the highly interesting internal factor of lexical innovation (borrowing and coining) due to word tabuing (Ahnne 1926, 1931; Henry 1927; Salmon 1927; Vernier 1948; White 1967 and 1968). As a consequence of the foregoing factors the Tahitian lexicon is considerably different today than it was two hundred years ago, probably having lost more words than it has gained.

WHY A NEW DICTIONARY? WHY NOW?

There never has been an orthographically adequate Tahitian-English dictionary. At a time when Pacific peoples are attaining greater independence and internal autonomy it is critical that outsiders have modern and useful tools for understanding and interaction with people of particular locales. The Francophone Pacific has been difficult of access to English speakers. An adequate dictionary will help solve this problem.

Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 343-353. *Pacific Linguistics*, C-88, 1985.

© Jack H. Ward

A complete (or relatively complete) dictionary of contemporary Tahitian may perhaps ignore archaic and obsolete words with little inconvenience for many users. There are, however, reasons for including such items in a full lexicon of any language. Properly noted older forms could shed light on the culture history of the speech community, its nature, origin, and subsequent interactions with other speech areas. It would also help address more general and theoretical questions of language change.

In the case of Tahitian there would be specific utility in a dictionary which would constitute a full record of the language. This is due to the growing role of Tahitian as it becomes a focus and symbol for increased cultural identity, self-awareness, and political action for most of the people in French Polynesia (Lavondes 1974; Tagapa 1979; Prevost 1970). Tahitian has been established as a subject of instruction in schools and competency in the language constitutes an option in testing for educational accomplishment and qualifying for local civil service. In 1978 Tahitian was established as a co-official language along with French and mechanisms for broad dialogue with the citizenry have been undertaken in order to help chart the course for development (LeMaitre 1978; Martin 1978; Pietri 1978; Salmon 1978; Tevane 1978).

A Tahitian Academy (*Fare Vana'a* in the vernacular), established by the Government in 1974, has had a great deal to do with the foregoing developments and it continues to investigate and promote the Tahitian language and culture, in part, by advocating broad language planning goals and by setting and pursuing the development of specific resource materials (Coppentrath 1975 and 1980). The Academy has established a more complete orthography than heretofore.

Current projects of the academy include a modern statement of Tahitian grammar, a new Tahitian-French dictionary, pedagogical materials for elementary schools, and the development of new vocabulary to meet the needs of Tahitian in a modern context. The grammar as well as the pedagogy and terminology projects have made substantial progress (*Fare Vana'a* 1978?; Service de l'Education et Académie Tahitienne 1979?; *Fare Vana'a* 1981 respectively).

The progress on a new Tahitian-French dictionary has necessarily not been as dramatic and a parallel Tahitian-English lexical project could serve several purposes, e.g. provide incentive, help develop methods and data, and possibly share in the distribution of certain tasks - all of which could benefit both efforts. Since one of the stated methods of the terminology project has been to utilise archaic Tahitian words (Coppentrath 1980), even without a concurrent Tahitian-French dictionary project, a clearer compilation of the Tahitian lexicon and its documentation through time could be of distinct utility. More generally such a work could provide linguistic data on the roots of and changes in Tahitian culture.

THE NATURE OF THE DICTIONARY

The lexical file is envisioned to be a dynamic data base. Publication of a dictionary would be only one of several outcomes. Due to the many input sources (below), the large volume of resulting data, the need to 'streamline' the file such as by purging it of redundant information, a multiplicity of uses (some of which may not yet be known), and the high probability that various sets and classes of data will need to be retrieved at different times it is clear that such an effort can only be handled economically and reliably through the

application of computer technology and, in this case, with the utilisation of programs now well developed at the University of Hawaii by Professor Robert Hsu.

The details of the program's design and capabilities are presented in Dr Hsu's paper (Hsu, in this volume). Here it is sufficient only to mention that a key aspect of his approach is the identification and labelling of the various distinctive categories of data in each entry of the various sources. The list of labels is unique to each research project but in any case is open-ended and can be augmented, reduced, reorganised, rearranged, and expanded as required. A complete and integrated file can be automatically maintained. The file and its data can be manipulated as needed. An English-Tahitian finderlist can be generated at will.

The elements (labels) presently being utilised or anticipated in the creation of the lexical file include the following.

ref:	Labels for:	Notes
a	Tahitian form	phonemically complete. Separate listings for homophonous forms.
	Sources *	Equatable to a chronology. All other labels the data of which derive from a given source will be identified by source
a	Orthography *	
a	Orthographic equivalents	where spelling differs greatly from a phonemic representation.
a	Phonetics	+ special phonetic features, e.g. forms which never show orthographic pronunciation.
a	Prosody	
a,b	Variant forms *	e.g. colloquialisms
a,b	Alternant forms *	
a,b	Cross reference *	
a,b	Synonyms *	
a,b	Antonyms	
a,b	Comparative *	i.e. cognates in other Polynesian languages
a,b,c	Protoforms	Reconstructed ancestral forms. Special reflexes of interest.
a ,c	Etymology	If not explained totally by the above or below.
c	Loanword *	Source language, lexeme, and process.
c	Folk etymology	
c	Change type	Loss, gain, loan translation, coinage.
c	Status *	e.g. rude, crude, swear word, euphemism.
c	Use *	Archaic, obsolete, rare, dialect, jargon, slang.

ref:	Labels for:	Notes (cont'd)
c	Word count	Frequency in text material such as early/modern literary, traditional tales, formal/colloquial speech.
d	Cultural domain	e.g. technology, plant, medicine ...
e	Classification *	Part of speech.
	-Example	(with English translation)
e	Reduplication *	Formal type and semantic class.
	-Example	
e	Morphological structure	
	-Example	
e,f	Semantic domain *	
f	Semantic range *	
f	English *	Subnumeration if needed for the delineation of the semantic range.
f	Literal meaning *	
f	Idiomatic meaning *	
f	French definition *	With English translation
f	Spanish definition *	With English translation
f	Latin names *	Scientific names
f	Place names *	Feature, location
f	Personal names	Proper, family, god
g	Author's comment *	
g	Note	All inputers may make comments, identify problems and ask questions. Subject of the note may be identified as phonology, grammar, morphology, history, etc.
g	Evaluation	on the reliability of information.
g	Tasking	Remaining work before completion
g	Review	

- * Identifies categories of information presently being extracted from the input sources.
a,b,c,d,e,f,g: See below

The contents of the entries within the file could follow the above order or be arranged in any other sequence desired. It may be noted however that the above order places the material in a sequence (sometimes overlapping) of broad categories of treatment. Thus information deals chronologically with (a) form, (b) reference to other forms, (c) evaluation, (d) social categories, (e) grammatical matters or the association of the entry with other forms within structure, (f) signification, and finally (g) house-keeping.

One is perhaps tempted to elaborate more categories than reasonable time will allow to be filled and utilised. However work can be done in stages and by more than one person. In any case a considerable amount of time and effort is required to categorise the content of several dictionaries and enter them into a file without error. It is the anticipation of increased accuracy and completeness as well as ultimate economies of time when the data is manipulated that motivates the enterprise.

THE SOURCES

The projected dictionary of Tahitian will be developed first from existing sources on the lexicon of the language. Later the file will be augmented from works such as ethnographies which incorporate language forms in their narratives. The first phase of development will be concluded with the examination of present-day written and spoken Tahitian materials already in collection. (Phase II which may overlap with part of phase I will be aimed at the computer-assisted consolidation and reduction of the file. Phase III is planned as the field investigation. Phase IV will be devoted to preparation of a dictionary.)

Within the first phase of development a number of sources are available which are historically distributed, albeit somewhat unevenly, over the two hundred and fifteen year span of time since the beginning of the written record of the language. Some works may have uncritically accepted forms from earlier works. The early wordlists are especially subject to inadequacies of transcription and definition. They can, however, establish a baseline, extend the time frame, and be very useful in revealing some language change which was in progress at the time of European advent. The chronological list below gives the sources presently known or suspected by the writer. * indicates sources that are presently being incorporated into the data base.

YEAR	IDENTIFICATION	SOURCE, if different
* 1768	Bougainville	LeMaitre
* 1769-	Cook	Lanyon-Orgill
* 1772-75	Spanish	Corney
1820-30?	Orsmond	?
* 1851	Davies	
* 1861	Jaussen	
* 1887	Jaussen	
* 1894	Jaussen	
1920-40s	Stimson	?
* 1939	Swift	
* 1944	Andrews	
1948	Jaussen	
* 1969	Jaussen	
* 1973	Cadousteau	
* 1973	LeMaitre	
* 1980	Fare Vana'a	

The use of several sources distributed over time will give the basic documentation for the rapid lexical change referred to above. When consolidated into a single file by computer the program will enable the file to reveal when words entered the record and when others were deleted. Deletion of words may be tentatively construed as definitive of obsolete status. Words appearing

later in the record may not however constitute new words in the language itself but merely represent more complete work by lexicographers. However the identification of some words as European loans, others as borrowings from Polynesian tongues, and still others as loan translations or coinages will help differentiate the pedigrees of the various forms. Changes in the semantic ranges of words will also be subject to investigation since the file will retain the chronology for all data including definitions. Studies in changes in orthographies and word classifications can also be made from the file.

In summary, the lexical files which will be developed may be used either for various investigations into the history of the Tahitian language or as a descriptive base into which may be incorporated additional and more complete information (especially as represented by the unasterisked categories of information on pages 345-346 above). For the present this paper is concerned only with the task of completing the composite lexical file.

RESULTING PROBLEMS AND STEPS TOWARD THEIR RESOLUTION

The incorporation of all Tahitian forms will create at least two sorts of problems, i.e. the identification of two classes of forms and the representation of the shapes of the non-contemporary ones. The first problem may be, in part, automatically handled by virtue of the ability to reveal each source of information and thereby the time periods from which the words are drawn. The matter could be handled graphically by using a different type face or the label 'obsolete' in the entries of words not in present day usage.

Care will have to be exercised and informants queried in order to reduce to a minimum the mislabelling of some words that have persisted as rare or dialect forms. The addition of word frequency counts can further assist users by revealing words of low utility. For example the word /māimoa/ *pet, plaything* is apparently a word of very low frequency since it does not appear in the 1973 LeMaitre general lexicon. Its occurrence and shape were verified by 1979 when the author noted it in the telling of a folktale recorded that year.

For words which continue to be classified as obsolete forms, and there will be a great many of them, the problem of shape is a very serious one. Phonologically distinctive vowel length is thoroughly provided only in the LeMaitre and Fare Vana'a lexicons (LeMaitre, 1973; Fare Vana'a, 1980). The latter work is quite limited in scope and only the former is designed for broad coverage. Since, by definition, obsolete words occur only in earlier sources all vowels are suspect. A more detailed study of the distribution of vowel length within words may reveal certain patterns that could reduce the likelihood that certain vowels would be long. At present however the writer is aware of very few such factors and offers some examples of vowel length occurrence.

1. Final vowel: /marō/ *dry*
/feiā/ *a class of people (not a social class)*
who share some characteristic
2. Penultimate: /mā'a/ *food*, /pāni/ *pan*
/piāna/ *piano*, /purōmu/ *road*
3. Final and penultimate: /mārō/ *stubborn, headstrong*
/tīmā/ *cement*, /pāpū/ *flat, level*

4. Antepenult: /p̄uai/ *strong*, /p̄arau/ *pearlshell*
 /p̄areu/ *cloth wrap around*
 /t̄orea/ *plover*, /t̄oro'a/ *profession*
 /p̄aruru/ *shield, protect*, /'ūputa/ *doorway*
5. Antepenult and final:
 /p̄aniē/ *a type of basket* (French *panier*)
6. Antepenult and penult:
 /t̄ī'ūe/ *to throw*
7. Antepenult, penult, and final:
 /t̄ōt̄ōvā/ and /t̄ōt̄ōā/ *harmful*
 /p̄ā'ō'ā/ *a type of dance*

Statistically type 2 is not highly frequent and types 5, 6 and 7 are of low frequency. However as can be seen all types occur. Partial reduplication tends to place length on initial vowels in some cases (/t̄a'oto/ *sleep*, /t̄ā'o'oto/ *sleep a little*, /t̄ā'oto'oto/ *sleep repeatedly*) and delete it in others (/p̄ā'i/ *to slap* and /papa'i/ *to slap*). Consequently very few uncertainties about vowel length are likely to be reduceable through more careful phonological analysis.

The glottal stop phoneme is most reliably represented in LeMaitre (1973) and Fare Vana'a (1980) but does occur in Andrews, Swift and the Drollet annotations of the Davies work. But care must be taken in all three sources, especially Andrews and Swift, and unless verifiable in LeMaitre or Fare Vana'a acceptance should be provisional. Drollet has not annotated all Davies forms, presumably because many forms were, in fact, obsolete by then. Some accents in Jaussen reveal glottal stop and the earlier editions were much more complete than the later editions where most of the accent marks were unfortunately deleted.

The glottal may logically occur between any two vowels. A combination of several early works may corroborate the presence of the glottal in some obsolete words.

The resolution of some cases may rest on external comparative evidence. Perhaps the most useful single tool will be the corpus of reconstructed Proto-Polynesian forms together with the cognate evidence (Biggs 1979). Dictionaries of individual languages will also need to be consulted. The glottal is perhaps more reliably and completely represented in dictionaries of other Polynesian languages than is the vowel length. However it is necessary to locate cognates in other languages whose regular sound correspondences with Tahitian are known.

Tahitian /ʔ/ frequently corresponds to /k/ and /ŋ/ elsewhere and these latter sounds are more likely to be transcribed properly even in earlier dictionaries. Tahitian /ʔ/ also reflexes PPN */k/* and */ŋ/*. Sometimes additional correspondences and reflexes must be taken into account in order to identify the proper corresponding forms in either related or ancestral languages.

Certain examples will illustrate this approach.

#	Tahitian (orthography)	Sources	Confirmed shape
1.	feau <i>hesitate</i>	Davies (da), Swift (sw) Jaussen (js), Andrews (an)	/feau/
	Basis = Maori wheau <i>stay, remain</i>		
2.	feraorao <i>daybreak</i> fera'ora'o	da, js an, sw, Drollet (dr)	/fēra'ora'o/
	Basis = Tuamotuan /herako/ <i>to become light as at dawn</i> Lengthening of first syllable in reduplications.		
3.	fetau <i>wrestle, combat</i>	da, js, an	/fētau/
	Basis = Rarotongan /'ētau/ <i>to frisk in sport, to make a pretence of a struggle</i>		
4.	fetoitoi <i>agitated as waves, sea:</i> feto'ito'i	da, js, Cadousteau (ca) sw	/fēto'ito'i/
	Basis = Tuamotuan /hētokitoki/ Vahitahi /fetokitoki/		

CONCLUSION

It is desirable to incorporate obsolete forms into a modern dictionary of Tahitian. Furthermore it is possible to make reliable estimates of the probable phonemic shape of such forms through the application of internal patterns of vowel length and external comparisons in the case of the glottal catch as well as vowel length.

BIBLIOGRAPHY

AHNNE, Édouard

- 1926 De la coutume du *pii* et des modifications qu'elle apporta au vocabulaire tahitien. *Bulletin de la Société d'études océaniques* 11:6-10. Papeete.
- 1931 La coutume du 'Pii' en Polynésie et le 'Fady' de Madagascar. *Bulletin de la Société d'études océaniques* 40/4(6):181-182.

ANDREWS, Edmund and Irene D. ANDREWS

- 1944 *A comparative dictionary of the Tahitian language*. Chicago: The Chicago Academy of Sciences. 253 pages.

BIGGS, Bruce

- 1979 *Proto-Polynesian Word List II*. Working Papers in Anthropology, Archaeology, Linguistics, Maori Studies No. 53. Auckland, N.Z.: Dept. of Anthropology, University of Auckland. 11 pages, + 2 microfiches.

CADOUSTEAU, Mai-Arii et J. ANISSON DU PERRON

- 1973 *Dictionnaire moderne français-tahitien et tahitien-français.*
Papeete, Tahiti: Stepolde. 336 pages.

COPPENRATH, Hubert

- 1975 *L'Académie Tahitienne.* Papeete: Te Fare Vana'a (from the *Journal de la Société des Océanistes*, numéro spécial, 264-301).
1980 *L'Académie Tahitienne après 6 ans d'existence.* *Bulletin de la Société d'études océaniques* 213/17 (2 Dec):762-771.

CORNEY, Bolton G.

- 1914 *The quest and occupation of Tahiti by emissaries of Spain during the years 1772-1776*, vol.2. London: The Hakluyt Society.

DAVIES, John

- 1851 *A Tahitian and English dictionary.* Tahiti: London Missionary Society Press. 314 + 7 pages.

FARE VANA'A

- 1978 *Précis de grammaire tahitienne.* MS. 96 pages.
1981 *Petit lexique du vocabulaire technique.* Papeete: Polytram 61 pages.

HENRY, Teuira

- 1928 *Ancient Tahiti.* *Bulletin* 48. Honolulu: Bishop Museum.

HSU, Robert

- 1985 *The role of computers in dictionary making.* In this volume, pp. 313-328.

JAUSSEN, Tepano

- 1961 *Grammaire et dictionnaire de la langue maorie: dialecte tahitien.* Saint-Germain-en-Laye: L. Toinon et Cie. 56, 78, 96 pp. (other editions are 1887, 1898, 1949, 1969).

LANYON-ORGILL, Peter A.

- 1979 *Captain Cook's South Sea Islands vocabularies.* London: Austin & Son.

LAVONDES, Henri

- 1974 *Language policy, language engineering and literacy in French Polynesia.* In Joshua A. Fishman, ed. *Advances in language planning*, pp. 255-276. (Revised from *Current trends in linguistics*, vol.8: *Linguistics in Oceania*, 1971:1110-1128).

LEMAÎTRE, Yves

- 1973 *Lexique du tahitien contemporain.* Paris: Office de la Recherche Scientifique et Technique Outre-Mer. 201 pp.
1978 *Social life and the promotion of Polynesian languages.* [translation] Paper delivered at the Colloquium on Polynesian Languages, Papeete, Tahiti.
1977 *Le vocabulaire tahitien de C.-F.-P. Fesche.* In Etienne Taillemite, ed. *Bougainville et ses compagnons autour du monde, 1766-1769*:166-177. Paris: Imprimerie Nationale.

LEVY, Robert

- 1970 Teaching of the Tahitian language in the schools of French Polynesia. In *Journal de la Société des Océanistes* 26:79-83.

MARTIN, John

- 1978 Use of "mass media" for the preservation and promotion of Polynesian languages. [translation] Paper delivered at the Colloquium on Polynesian Languages, Papeete, Tahiti.

ORSMOND, John Muggridge

- n.d. The Ariori - war in Tahiti, in Orsmond papers 1816-1892?, vol.4. (MLA 2608, also microfilm CY666); part of Tahitian dictionary, in Orsmond papers, vol.5. (MLA2609). Mitchell Library, Sydney.

PIETRI, Raymond

- 1978 An evaluation of studies and actions taken for the preservation and promotion of Polynesian languages. [Translation] Paper delivered at the Colloquium on Polynesian Languages, Papeete, Tahiti.

PREVOST, Paul J.C.

- 1970 L'expansion de la zone d'influence de la langue tahitienne. *Journal de la Société des Océanistes* 26(Sept):256-260.

SALMON, Marau Taaroa

- 1927 Notes et commentaires sur le Pii: attributes royaux ... *Bulletin de la Société d'études océaniques*. 20:260-271.

SALMON, Nedo

- 1978 Work plan for the preservation and the promotion of Polynesian Languages. [translation] Paper delivered at the Colloquium on Polynesian Languages, Papeete, Tahiti.

SERVICE DE L'EDUCATION ET ACADEMIE TAHITIENNE

- 1979? *Ta'u Puta Reo Tahiti*. 4 vols. Papeete.

STIMSON, John Frank

- n.d. Tahitian lexical file. MS. (uncertain as to existence and location.)

SWIFT, Edward

- 1939 A dictionary of the Tahitian language. Papeete: Manuscript for Eastham Guild. 542 pages.

TAGAPA, William E. H.

- 1979 The Tahitian language: a historical and vernacular controversy. *Pacific Studies* 2/2:144-155.

TEVANE, Maco

- 1978 The present situation of Polynesian languages within French Polynesia. [translation] Paper delivered at the Colloquium on Polynesian Languages, Papeete, Tahiti.

VERNIER, Charles

- 1948 Les variations du vocabulaire tahitien avant et après les contacts européens. In *Journal de la Société des Océanistes* 4:57-85.

WHITE, Ralph Gardner

- 1967 Onomastically induced word replacement in Tahitian. In *Polynesian culture history: essays in honor of Kenneth P. Emory*, 594 pp. B.P. Bishop Museum Special Publication 56. Honolulu: Bishop Museum Press.
- 1968 Borrowing and taboo in Eastern Polynesia. *Journal of the Polynesian Society* 77/1:57-76.

