

Trends in Linguistics Documentation 10

Editors

Werner Winter
Richard A. Rhodes

Mouton de Gruyter
Berlin · New York

Comparative Austronesian Dictionary An Introduction to Austronesian Studies Part 1: Fascicle 1

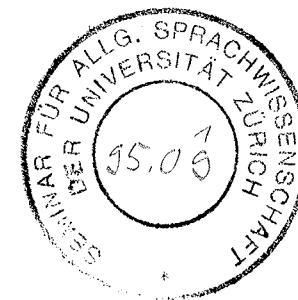
edited by
Darrell T. Tryon

Assistant Editors
Malcolm D. Ross
Charles E. Grimes
Adrian Clynes
K. A. Adelaar

Editorial Assistant
Ling Matsay

Area Coordinators

Taiwan	Shigeru Tsuchida
Philippines	R. David Zorc
Indonesia	Anton Moeliono
Papua New Guinea	Malcolm D. Ross
Solomons/Vanuatu	Darrell T. Tryon
New Caledonia	Jean-Claude Rivierre
Fiji/Polyynesia	Paul Geraghty
Micronesia	Sheldon P. Harrison



Mouton de Gruyter
Berlin · New York 1995

© Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication-Data

Comparative Austronesian dictionary : an introduction to
Austronesian studies / edited by Darrell T. Tryon.
p. cm. — (Trends in linguistics. Documentation ;
10)
Includes bibliographical references and index.
ISBN 3-11-012729-6 (alk paper) :
1. Austronesian languages — Synonyms and antonyms —
Dictionaries. 2. Austronesian languages — Etymology —
Dictionaries. 3. Austronesian languages — Glossaries,
vocabularies, etc. 4. Malayan languages — Glossaries,
vocabularies, etc. 5. Asia, Southeastern — Languages —
Glossaries, vocabularies, etc. I. Tryon, D. T. (Darrell T.)
II. Series.
PL5026.C66 1994
499'.2—dc20
94-20164
CIP

Die Deutsche Bibliothek — Cataloging-in-Publication-Data

Comparative Austronesian dictionary : an introduction to
Austronesian studies / ed. by Darrell T. Tryon. — Berlin ; New
York : Mouton de Gruyter.
(Trends in linguistics : Documentation ; 10)
ISBN 3-11-012729-6
NE: Tryon, Darrell T. [Hrsg.]; Trends in linguistics / Documen-
tation
Pt. 1 Fasc. 1 (1994)

© Copyright 1994 by Walter de Gruyter & Co., D-10785 Berlin
All rights reserved, including those of translation into foreign languages. No part of this
book may be reproduced or transmitted in any form or by any means, electronic or mechani-
cal, including photocopy, recording or any information storage and retrieval system, with-
out permission in writing from the publisher.
Printing: Ratzlow-Druck, Berlin.
Binding: Lüderitz & Bauer, Berlin.
Printed in Germany.

Table of contents

List of contributors	ix
List of abbreviations	xiii
Language abbreviations	xvii
Table of phonetic symbols	xviii
List of maps	xxi
Acknowledgments	xxiii
1. Introduction to the comparative Austronesian dictionary	Darrell T. Tryon 1
2. The Austronesian languages	Darrell T. Tryon 5
3. Some current issues in Austronesian linguistics	Malcolm Ross 45
4. Listing of Austronesian languages	Barbara F. Grimes — Joseph E. Grimes — Malcolm Ross — Charles E. Grimes — Darrell T. Tryon 121
5. Introductions to individual languages	281
1. Atayal	Paul Jen-kuei Li 283
2. Tsou	Shigeru Tsuchida 293
3. Rukai	Paul Jen-kuei Li 297
4. Paiwan	Dah-an Ho 307
5. Yami	Anne West 315
6. Isnag	Rodolpho Barlaan 321
7. Kalinga Limos	Harmut Wiens 329
8. Tagalog	R. David Zorc 335
9. Aklanon	R. David Zorc 343
10. Palawan	Nicole Revel 351
11. Molbog	R. David Zorc-Arnold Thiessen 359
12. Kagayanan	Louise A. MacGregor 363
13. Sarangani Blaan	Mary E. Rhea 369
14. Bāngingi Sama	Eunice Diment 375
15. Timugon Murut	D. J. Prentice 381
16. Malagasy	K. A. Adelaar 393
17. Aceh	Mark Durie 407
18. Toba Batak	K. A. Adelaar 421
19. Minangkabau	K. A. Adelaar 433

- Teeuw, A. – H.W. Emanuels
 1961 *A critical survey of studies on Malay and Bahasa Indonesia* (KITLV 5) (The Hague: Nijhoff).
- Tuuk, H.N. van der
 1861 *Bataksch-Nederduitsch woordenboek* [Batak-Dutch dictionary] (Amsterdam).
 1864 *Outlines of grammar of Malagasy language (Journal of the Royal Asiatic Society 8/2).*
 1864, 1867 *Tobasche spraakkunst.* [A grammar of Toba-Batak] (2 parts) (Amsterdam: Muller).
- Wurm, Stephen A.
 1972 *The languages of Australia and Tasmania* (The Hague: Mouton).
 1982 *The Papuan languages of Oceania* (Tübingen: Gunter Narr).
- Wurm, Stephen A. – Lois Carrington (eds.)
 1978 *Second International Conference on Austronesian Linguistics: proceedings* (2 fascicles) (*PL*, C-61).
- Wurm, Stephen A. – Shirô Hattori (eds.)
 1981 *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).
 1983 *Language atlas of the Pacific area, part II: Japan area, Taiwan (Formosa), Philippines, mainland and insular south-east Asia* (Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy; *PL*, C-67).
- Wurm, Stephen A. – Donald. C. Laycock (eds.)
 1970 *Pacific linguistic studies in honour of Arthur Capell* (*PL*, C-13).

3. Some current issues in Austronesian linguistics

Malcolm D. Ross

Introduction

A brief history of Austronesian studies and a thumbnail sketch of some current opinions concerning the genetic classification of Austronesian languages is provided by Tryon (this volume). A major purpose of this chapter is to provide for readers outside the narrow confines of Austronesian linguistic prehistory some indication of the strengths and weaknesses in our current knowledge, to point to possible directions of development in this understanding, and to indicate where fresh research is especially necessary. The issues to be reviewed here concern the reconstruction of Proto-Austronesian phonology and morphosyntax and the internal and external genetic relationships of Austronesian languages.¹

Subgrouping methodology

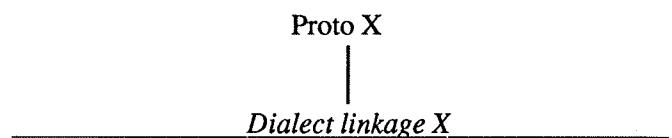
The tree diagrams in Figures 1 to 8 of the previous chapter provide a sketch of our present understanding of the subgrouping of Austronesian languages. However, these diagrams entail a certain ambiguity, as they fail to distinguish between two different ways in which Austronesian languages and subgroups have diverged from each other: some have diverged by *dialect differentiation*, others by *separation*.²

Divergence by *dialect differentiation* has occurred when the dialects of a language spoken in different areas of the speech community have gradually become more and more different from each other, until the differences are so great that we must speak of different languages rather than of dialects. The process of differentiation has often been accelerated by the fact that among Austronesian speakers in the past the most important unit of political organisation was typically the village, and people have tended to identify with their village rather than with other speakers of their language. Peculiarities of a village's speech have tended to become an emblem of village identity both for the speakers and their neighbours, and as a result

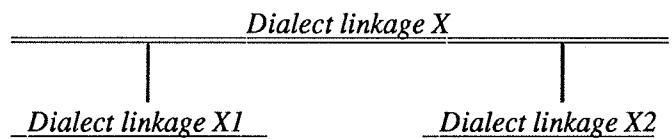
these peculiarities have been emphasised by their speakers but shunned by their neighbours. We may conveniently use the term *linkage* to refer to a group of languages which have arisen by dialect differentiation.

As the term implies, languages diverge by *separation* when two or more communities speaking the same language become sharply separated socially and/or geographically so that contact between them is more or less severed and as a result the two languages change in different ways and perhaps at different speeds.

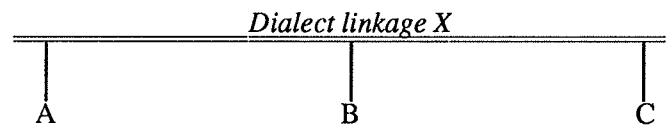
Diagrammatically, the fact that a language Proto-X has differentiated into a dialect linkage may be represented by a double line:



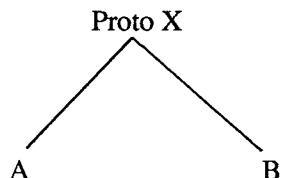
Sometimes, a linkage splits into two (or more) linkages:



Or into separate languages, here labelled A, B and C:



On the other hand, the fact that languages A and B are descended from the language Proto X by separation can be indicated by a branching node:



Whether languages diverge by dialect differentiation or by separation, it is clear that languages which belong to the same subgroup must be descended from a common ancestor at some time-depth or other. Thus in Figure 1, which shows the primary branches of the Austronesian language family, we could be pedantic and label the top node "Proto-Austronesian" instead of "Austronesian", since in linguistic terminology a putative earlier language is labelled with the prefix "Proto-", whereas an unprefixed term like "Austronesian" is taken to refer to a group of languages. At this point, however, a difficulty arises. Consistency would seem to require that we also label the branches "Proto-Atayalic", "Proto-Tsouic", "Proto-Paiwanic" and "Proto-Malayo-Polynesian", rather than simply "Atayalic", "Tsouic" and so on. But we can be reasonably confident that Proto-Austronesian (henceforth PAN) was spoken on Taiwan (see previous chapter; also Blust 1985, Bellwood 1985), and that the languages of the Atayalic, Tsouic and Paiwanic groups have had a continuous history of at least five thousand years on that island. It is therefore plausible that the languages of the Atayalic, Tsouic and Paiwanic groups have arisen by gradual dialect differentiation from the PAN (or its early descendant dialects) of those who stayed behind. This could mean that there never was a "Proto-Atayalic", a "Proto-Tsouic" or a "Proto-Paiwanic".

If we were to attempt to redraw Figure 1 less ambiguously using the conventions above, we might arrive at something like Figure 9. This says that Proto-Malayo-Polynesian (henceforth PMP) diverged from the Austronesian languages of Taiwan as the result of separation (when its speakers left Taiwan more than 4,500 years ago). This is a reasonable assumption. Figure 9 also says that the languages of the Atayalic, Tsouic and Paiwanic groups are descended, not from separate proto-languages, but from three dialect linkages which resulted from the break-up of the Formosan dialect linkage³ into which PAN had differentiated around the time of the departure of PMP speakers. Whilst this is a plausible sequence of events, it is by no means the only possible one. There is some evidence, for example, that Atayalic forms one subgroup of languages and all the other Austronesian languages on Taiwan another, and that this latter subgroup is then divisible into the Tsouic and Paiwanic (and perhaps other) groups. However, even if this is so, we do not know (a) whether the primary split into Atayalic and non-Atayalic took place before or after or at about the same time as the departure of the speakers of PMP, and (b) whether these divisions took place by dialect differentiation or by separation, i.e. whether there ever was a "Proto-Atayalic", a "Proto-Tsouic" or a "Proto-Paiwanic". (Available evidence makes Proto-Paiwanic less likely than the other two.)

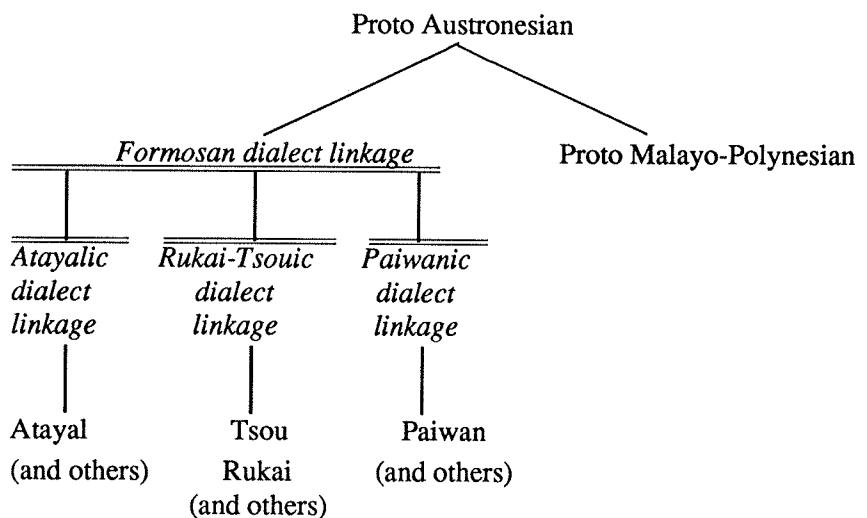


Figure 9. A hypothetical (but not necessarily correct) representation of the earliest branchings of the Austronesian language family

The point of this discussion is not to debate the early diversification of Austronesian languages, but rather to show that figures 1-8 are ambiguous — and that this ambiguity is a virtue, since to be more precise would imply that we know considerably more than we actually do.

We turn now to the question of how we can know what subgroups languages belong to and whether they have diverged by dialect differentiation or by separation.

Two methods are commonly used to determine the genetic relationships of languages and thereby to combine them into subgroups. These are the “classical” *comparative method* and the method of *lexicostatistics*.

Under the comparative method, a set of related languages is systematically compared and the proto-language from which they are supposedly descended is reconstructed. The languages in the set are then examined to determine what innovations in phonology and morphosyntax each reflects relative to the proto-language. These innovations typically pattern across languages in one of two ways, indicating whether the languages have diversified by separation or by dialect differentiation. When separation has occurred, groups of languages share discrete bundles of innovations. For example, within a set of languages A to Z whose reconstructed proto-language is Proto-AZ, languages A to M share a given

bundle of innovations (which are not shared by languages N to Z). From this we infer that A to M are descended from Proto-AM (itself a descendant of Proto-AZ) and that they share this bundle of innovations because the innovations occurred in Proto-AM itself.

In the second patterning of innovations, which characterises divergence by dialect differentiation, shared innovations form overlapping, rather than discrete, bunches. For example, among a set of languages A-G scattered in a line along a coastline, A, B and C may reflect one bunch of innovations, C, D, and E another bunch, and D, E, F, and G yet another bunch: that is, the bunches overlap in chain-like fashion. In an inland area or an archipelago, overlapping bunches of innovations may form a more complex network-like pattern. We might find innovations shared, for example, by A, B, and C, by C, D, E, and F, and by A, D, F, and G. Such overlaps arise because at some stage in the past proto-language AG has differentiated out into a dialect linkage (i.e. a chain or a network) whose member-dialects have remained in sufficient contact with each other for an innovation generated in one dialect of the linkage to diffuse through adjacent dialects of the linkage, i.e. undetected early borrowing has occurred.

There is one set of elements in language change which the classical comparative method does not readily recognise. Contact-induced change (change brought about in a dialect or language through its speakers' contact with a neighbouring dialect or language) does not only occur when neighbouring languages are closely related; it may also occur when they are genetically only distantly related or completely unrelated. For example, there are Austronesian languages in Melanesia whose grammars have undergone wholesale modification in the direction of neighbouring Papuan (non-AN) languages. Their speakers have become so much at home in the language of their neighbours that they have reshaped their language along Papuan lines. Interestingly, the basic vocabulary of such languages often includes few borrowings, because it is in the words themselves (and not in the structures in which the words are used) that speakers perceive the essence of the language and therefore the emblems of their own identity. Although the importance of this kind of change has been recognised for eighty years (see Ray 1911, 1926), it has received only sporadic attention from linguists working in Austronesian languages (Capell 1943, 1976, Lynch 1981, Pallesen 1985, Ross 1987) and is a major research need.

There is, however, a second, rather different kind of contact-induced change where vocabulary is borrowed, sometimes on quite a large scale, from a culturally dominant language, and the word lists note many such borrowings: from Sanskrit in western Indonesia (as a result of Indian cultural

influence), from Arabic in the Islamicised sections of the region, from Malay throughout Indonesia and the Philippines (through its long-time use as a *lingua franca*),⁴ and throughout the Austronesian region from the languages of colonial powers (Chinese and Japanese in Formosan languages, Spanish and English in the Philippines, Dutch in Indonesia, Spanish and Japanese in Micronesia, and French and English in the south Pacific islands). Borrowing of this kind has also occurred in small-scale societies, as Dutton (1982a) has documented for a south-east Papuan case.

A third kind of contact-induced change occurs when, after a period of bilingualism, a group changes its first language, and features of the former first language are retained in the adopted language, for example, in the way they pronounce it.⁵

Contact-induced change also causes problems for the second method referred to earlier, lexicostatistics. The procedure in lexicostatistics is to collect a list of words for the same set of meanings from a number of languages, and for every pair of languages marking each pair of words as either "cognate" (i.e., the words with the same meaning in the two languages are deemed to be related) or "non-cognate", and calculating the percentage of the items in the list which are deemed cognate for each pair of languages. This "cognate percentage" is then taken to be a measure of the degree of genetic relationship between the two languages: the higher the percentage, the closer the assumed relationship.

Apart from the difficulty of deciding whether words of similar meaning in two languages are in fact cognate,⁶ the lexicostatistical method encounters the problem that words, like other features of language, may be subject to contact-induced change, that is, they may be borrowed by one language from its neighbour. Items of vocabulary will readily pass between two Austronesian languages which are not necessarily each other's closest relatives, and thereby distort a lexicostatistical classification.

This difficulty is manifest when the comparative and lexicostatistical methods are applied to the same group of languages and give conflicting results. In a well-known study, Dyen (1965b) used lexicostatistics to derive a genetic tree of the Austronesian language family. However, a number of the groupings which Dyen proposed have not stood up under the magnifying glass of the comparative method.⁷ Since the classical comparative method sets out to reconstruct the prehistory of the morphemes of a set of languages as well as their divergences, whereas the lexicostatistical method reconstructs only the history of language splits, the former method, complicated and time-consuming as it may be, is clearly preferable to the

latter, and it is therefore important to ask with regard to any classification of languages which method it is based on.

An extension of lexicostatistics known as "glottochronology" has also been applied in Austronesian studies. Glottochronology is based on the assumption that words associated with the non-cultural meanings in a lexicostatistical list are replaced at a constant rate over time, and that the lexicostatistical percentage of shared cognates for a pair of languages can therefore be used not only to ascertain the genetic relationships between them, but also to determine the date at which their parent language split into the two daughter-languages. However, if the lexicostatistical method is unreliable, the associated glottochronological method must also be unreliable (see Grace 1966). Blust (1981b) has demonstrated on the basis of a study of Austronesian languages that words are *not* replaced at a constant rate over time. Bellwood (1978: 125-126), seeking to date the break-up of Proto Oceanic (the language ancestral to all the Austronesian languages of Oceania) found a puzzling discrepancy between the glottochronological dating of its break-up to no later than 3000 B.C. and the archaeological dating of Austronesian-related artefacts from the area of the Proto Oceanic homeland in western Melanesia, for which the earliest dates were only about 1500 B.C. The discrepancy is at least partly attributable to the unreliability of glottochronology.

In the comparative method, reconstruction and subgrouping are interdependent. A prerequisite for the tenability of any subgrouping hypothesis must be an adequate reconstruction of the proto language: subgrouping under the comparative method is based on innovations which have occurred in present-day languages relative to a proto language. If the reconstruction of the proto language is inadequate, then wrong inferences are likely to be made about innovations and consequently about subgrouping. Conversely, however, decisions about reconstruction are dependent on hypotheses about subgrouping.

An example will serve to illustrate this interdependence. Phonemes in a proto language are reconstructed from correspondences among sounds in the languages descended from it. In Table 1, cognate sets for six items are given across twelve Austronesian languages. Atayal, Tsou, Rukai, Paiwan, Amis and Bunun are all spoken in Taiwan; Tagalog in southern Luzon (Philippines); Toba Batak in north-western Sumatra (western Indonesia); Uma on Sulawesi (north central Indonesia); Manggarai on Flores (south central Indonesia); Kairiru on the island of the same name off the north coast of Papua New Guinea; and Samoan in the central Pacific. The words listed in Table 1 illustrate two sound correspondences. The words for 'ear', 'eye'

Table 1. Cognate sets reflecting PAN *t and *C⁸

	ear	eye	head louse	three	freshwater eel	seven
Atayal (Atayalic)	čanja?	-	kuču?	tu-yat	tuča-kiy	ma-pitu?
Tsou (Rukai-Tsouic)	-	mɸō	kɸū	turu	tuŋ-roza	pitu
Rukai (Rukai-Tsouic)	ɸaliŋa	maɸa	koɸo	tolo	tola	pito
Paiwan (Paiwanic)	ɸaliŋa	maɸa	kɸəlū	cəlū	cul'a	picu
Amis (Paiwanic)	talŋa?	mata?	kutu?	tulu	tuɬa	pitu
Bunun (Paiwanic)	taiŋa	mata?	kutu?	tau?	-	pitu?
Tagalog (WMP)	tēŋa	mata	kuto	ta-tlo	-	pito
Toba Batak (WMP)	-	mata	hutu	tolu	-	pitu
Uma (WMP)	tiliŋa	mata	kutu	tolo	-	pitu
Manggarai (CMP)	-	mata	hutu	təlu	tuna	pitu
Kairiru (Oceanic)	tiliŋe	mata	κut	tuol	tun	-
Samoan (Oceanic)	taliŋa	mata	?utu	tolu	tuna	fitu
PAN	*Calŋa	*maCá	*kúCuh	*telúh	*tuLá	*pitú
PMP	*talŋa	*matá	*kútuh	*telúh	*tuná	*pitú

and ‘head louse’ exemplify the correspondence between č in Atayal, φ in Tsou, Rukai and Paiwan, and t in the other eight languages. The words for ‘three’, ‘freshwater eel’ and ‘seven’ exemplify a second correspondence, between c in Paiwan and t in all the other listed languages. From these two sound correspondences, the two PAN proto-phonemes *C and *t are conventionally reconstructed, and we can attribute to them the probable phonetic values [ɸ] and [t].

This reconstruction both suggests and entails a subgrouping hypothesis. The lower eight languages in our sample all have t where Atayal has either č or t and Tsou, Rukai and Paiwan have either φ or t; that is, they share in the merger of the PAN proto-phonemes *t and *C. This suggests a hypothesis that these eight languages belong to a subgroup (but see below): they share a common ancestor in which the merger of *t and *C had taken place. But note that the reconstruction itself already entails such a hypothesis. An alternative hypothesis would be that PAN had only the proto-phoneme *t, which split in a common ancestor of Atayal, Tsou, Rukai and Paiwan into *t and *C. Whichever reconstruction is adopted, it entails an implicit subgrouping decision.

How can we escape this circularity? In the case of PAN,⁹ we are dependent on evidence based on likely directions of phonetic change. Is it more likely that PAN *C [ɸ] and *t [t] merged as [t], or that PAN *t [t] split

into [ɸ] and [t]? A split would be the more likely hypothesis if it could be shown that [ɸ] was the outcome before high or front vowels and [t] elsewhere, since such splits are quite common in the world’s languages. But the data show no such conditioning, leaving the merger as the more likely direction of change.¹⁰

In practice, decisions of this kind are rarely dependent on a single innovation. In the present case, the weight of evidence, both phonological and morphosyntactic, suggests that several innovations besides the merger of *C and *t should be attributed to the lower six languages of Table 1 and to all Austronesian languages outside Taiwan and that these languages form a subgroup which is labelled “Malayo-Polynesian” (Figure 1).¹¹ This decision leaves no innovations shared by all the languages on Taiwan, implying that they constitute several first-order subgroups of Austronesian. It is this decision which informs the reconstruction of PAN phonology below.

This example illustrates how reconstructing something of the phonetics of the proto-language is vital to subgrouping procedure. PAN *C has been written thus by most scholars since Dyen (1965a), and there has been a tendency among Austronesian scholars (and among comparative linguists in general) to treat reconstructed proto-phonemes as algebraic cover symbols of unknown phonetic value, such that the actual choice of symbol is unimportant.¹² However, a consequence of this convention is that identifying the direction of phonetic change becomes impossible, and we have no means of knowing whether *C > t or *C > ts is an innovation, because the phonetic value of *C is unknown. For example, Li (1985) adduces five shared innovations to show that the languages of the Atayalic group and several languages placed by other scholars in the Paiwanic group on the contrary share a common ancestor and therefore form a Northern Formosan subgroup. One of these innovations is that PAN *C becomes ts or s. But as we have seen, this is probably not an innovation: the innovation occurs instead in those languages where *C ([ɸ]) has become t.

The “algebraic” tendency in PAN reconstruction is rooted in its history. When Dempwolff (1934, 1937, 1938) initially reconstructed his *Ur-Austronesisch* (‘Proto-Austronesian’), he used no Formosan data. As a result, what he reconstructed was what we would now call PMP. When Formosan data became available to Dyen, he added new proto-phonemes to Dempwolff’s reconstructed system (see Dyen 1965a) to take account of his new insights. However, with these and other additions, the conventionally accepted PAN phoneme system became overloaded with symbols to which it was no longer possible to attach phonetic values.¹³ Various scholars (Dahl

Table 2. Approximate equivalences among reconstructed phoneme systems¹⁴

Dempwolff (PMP)	Dyen (PAN)	Tsuchida (PAN)	Dahl (PAN)	Ross (PAN)
p	p	p	p	p
b	b	b	b	b
m	m	m	m	m
v	w	w, W	u	w
t	t	t	t ₁	t
ʈ	T	T
...	C	C	t ₂	C
d	d	{D ₃ , d- D ₂ , -d D ₁ }	d ₁ d ₂ d ₃	d ₁ d ₂ d ₃
ɖ	D	D
n	n	n	n	n
...	s ₂	θ	θ	...
t'	s ₁	s	t'	s
d'	z	z	d'	...
...	Z	Z, d-	Z	Z
n'	ñ	ñ	n'	...
l	l	l	l	l
...	L/N	N	ɿ	L
ɿ	r	r	r	r
...	S, X, x	S, X, x	S	S
j	y	y	i	y
k'	c	...	k'	...
k	k	k	k	k
g	g	g	g	g
g'	j	j	g'	j
ɳ	ɳ	ɳ	ɳ	ɳ
ɣ	R	R	ɣ	R
h	q	q, Q	q	q
...	h	h	H	h
...	?	?	...	?
i	i	i	i	i
u	u	u	u	u
ə	e	ə	ə	e
a	a	a	a	a

1973, 1981; Harvey 1982; Wolff 1974a, 1982, 1988) have protested at this overloading, and have attempted to reconstruct a plausible PAN phoneme system.

A second reason for the continued use of the “algebraic” notation is that it has come to be used for both PAN and PMP, and its use for both proto-languages facilitates comparison. However, if the phonological system of PMP differed significantly from that of PAN (as seems to be the case), then the use of the same orthography for both proto-languages is somewhat misleading.

Reconstructing Proto-Austronesian phonology

Because of the importance of attributing phonetic values to the symbols used in reconstruction, this section is devoted to a presentation of these symbols and a discussion of their likely values. The views presented here differ in some respects from those of other scholars, and are certainly not the last word on the PAN phonology; on the contrary, they are presented in order to show that unresolved issues remain in this connection. A detailed account of academic debate about PAN phonology would require several chapters to itself. The issues under debate are discussed by Dahl (1973, 1976, 1981), Wolff (1974a, 1982, 1988), and in Blust’s (1976) review of Dahl (1973). Blust (1990b) includes a brief overview of the PAN phonological system and its putative phonetics. Ross (1991b) provides a more detailed account of some of the matters which receive comment here. There are a number of orthographic conventions used in the literature, and Table 2 provides equivalences between the orthographies of Dempwolff, Dyen (also used by Blust, Zorc and many other scholars), Tsuchida (1976), Dahl, and that adopted in the present discussion. Dyen’s system is the most widely used orthography, and provides the basis of the latter.

In attributing values to the symbols of the PAN orthography, it is assumed here that the languages of Taiwan represent several first-order subgroups of Austronesian and, as illustrated above, are of considerable importance for reconstructing PAN. Distinctions which Dempwolff reconstructed, but which do not occur in Formosan languages, are taken to be innovations which occurred in PMP (or later) and are not reconstructed for PAN. By the same token, distinctions not reconstructed by Dempwolff but warranted by the Formosan data are reconstructed here.¹⁵ In Table 3, which summarises the data on which our interpretation of PAN phonology is based, these distinctions are marked by adding subscripts to Dyen’s orthography in the leftmost column.

Table 3. Formosan evidence for the reconstruction of PAN phonemes¹⁶

PAN	Proto-Atayal	Saisiyat	Bunun	Thao	Saaroa	Proto-Rukai	Proto-Puyuma	Proto-Paiwan
p	p	p	p	p	p	p	p	p
b	b	β	b	ϕ	β	b	b [b β]	β
m	m	m	m	m	m	m	m	m
w	w	w, ?	v	w	∅	w	w	β
t	t	ʈ	t	ʈ	t	t	t	c
C	l/_(V)N č-č-t	ʂ	t	θ	ɸ	ɸ	ʈ	ɸ
Z	d	r	d	ʂ	s, ɸ	d	d [d ɖ]	ʐ
d ₁	d	r	d	ʂ	s, ɸ	d	d [d ɖ]	ʐ
d ₂	d	r	d	ʂ	s	ɖ	ɖ [ɖ, z]	z
d ₃	d	r	d	ʈʂ	s	ɖ	ɖ [ɖ, z]	ɖ
L	ʈ	ɿ	n	ɖ[ɖ, ʈ]	ʈ	l	l, n	l ^y
l	r[r, ŋ]	l[l, l ^y]	∅	ř	ř	l	ř	l
r	ř	ř	ř	ř	ř
S	ʂ, h[h]	ʂ	s	ʂ, ∅	s, ∅	s	∅	s
s	h, x	h	č	ʈ	s, ∅	θ	s	t
j	∅/_ə; y	z	∅	ɖ	∅/i; ʈ	∅/_i; g	d [d, ɖ]	d
n	n	ɳ	n	n	n	n	n	n
y	y	y	ɖ	y	ʈ	-ɖ-y	y	y
k	k	k	k	k	k	k	k	k
g	k	...	k	...	g [g, h]	g
ŋ	ŋ	ŋ	ŋ	n	ŋ	ŋ	ŋ	ŋ
q	?	k	k	?, ∅	∅	? [? ɸ]	k	?
h	h[h]	h	h-∅	∅	∅	∅	∅	∅
?	∅	∅	?	∅	∅	∅	∅	∅
R	r/_i; g	l	l	l	r	r, ?	r[ř]	∅
i	i	i	i	i	i	i	i	i
u	u	œ/*q,s	u	u	o	u	u	u
a	a	æ/*q,s	a	a	a	a	a	a
e	ə,u	i/*q,s	i, u, a	i, u	ə	ə	ə	ə

Table 4. Possible phonetic values of the reconstructed Proto-Austronesian phoneme system

	bi-labial	dental/alveolar	alveolar affricate	retroflex	palatal	velar	uvular	glottal
stop vl	p	t	C [ɸ]			k	q [k]	?
stop vd	b	d ₁ [d]	d ₂ [dz]	d ₃ [ɖ]	Z[j]	g		
nasal	m	n				ŋ		
fricative vl			s		S [ʂ, ʂ]			h
fricative vd			j[z]					
lateral		L [l, ɿ]			l [l, ɿ]			
trills		r [ř]					R	
semivowels	w	y						
	i		u		-iw			-uy
		e [ə]						
		a					-ay	-aw

Over the last twenty years scholars working in Formosan languages have done a great deal of careful comparative work, reconstructing proto-languages ancestral to the various Formosan language groups, and these largely provide the basis of the material in Table 3. Before this work became available, major difficulties in using Formosan data had been that the latter often contained errors, data from different dialects of a language were mixed, and the phonologies of individual languages were poorly understood (this is true of some of the material in Dahl 1973). The collection of comparative data and the reconstruction of lower-order proto-languages has greatly enhanced our understanding of Formosan languages, and the data underlying Table 3 are accessible in published form. Where relevant, these data have also been cross-checked against the lists in the present work.

For comparison's sake, the orthographies of the various sources represented in Table 3 have been replaced by a standard representation of their phonetic values. Two unbracketed symbols separated by a comma represent an apparent phonemic split: e.g. PAN *S split in Proto-Atayal into *ʂ and *h. Two symbols in square brackets represent allophones of the phoneme preceding the brackets: e.g. Thao ɖ has the allophones [ɖ] and [ʈ]. The bilabials and velars of the PAN phonological system shown in Table 4 are uncontroversial, except for *g. Wolff (1973, 1982) has suggested that *g

and *r not be reconstructed. However, in any language there are phonemes which are used much less often than others, and it seems that the PAN consonants with lower functional load were *g, *r and *j. Certainly, Formosan data require their reconstruction, and one item, *garu ϕ ‘comb’, contains both *g and *r.¹⁷

The dental/alveolar, alveolar affricate, palatal and retroflex — collectively the coronal — proto-phonemes are most easily discussed together. There is good evidence for the reconstruction of a contrast between dental/alveolar and retroflex coronals, at least in the cases of the voiced stops and the liquids. Note, however, that the phonemes to which a retroflex value is attributed here are not those which have been conventionally reconstructed as retroflex. At the same time, we find no direct Formosan evidence for the reconstruction of the conventionally recognised palatal consonants. Apart from the probable palatal *Z, the conventionally reconstructed retroflexes and palatals are reflected only in extra-Formosan languages, and are therefore more appropriately reconstructed — if at all — either within the phonological system of PMP or in a descendant thereof. We return to this matter below.

The voiceless dental/alveolars *t and *C were illustrated and discussed above. Dahl (1976: 55-60) proposes that the Formosan data require the reconstruction of three voiced dental/alveolars, namely *d₁ [d], *d₂ [dz] and *d₃ [d̪]. The data in Table 3 support these with consistent correspondences among Formosan languages. Whilst only Proto-Paiwan has this three-way distinction (reflected in some present-day Paiwan dialects; Ho 1978), the contrast between PAN *d₁ on the one hand and PAN *d₂ and *d₃ on the other is also reflected in Proto-Puyuma. The Formosan reflexes of *Z are identical to those of *d₁, except in Proto-Rukai, where *Z is reflected as alveolar /d/ and *d₁ merges with *d₂ and *d₃ as retroflex /d̪/. However, as Tsuchida (1976) and Dahl (1981: 98) note, correspondence between Proto-Rukai /d/ and PMP *Z requires the reconstruction of PAN *Z, probably a palatal stop or affricate. These matters are examined in more detail in Ross (1991b).

The phonetic values attributed to PAN *d₁, *d₂ and *d₃ are based on two factors. Firstly, the distinction between dental/alveolar and retroflex coronals is clear across a substantial portion of Taiwan (see Table 3), accounting for *d₁ and *d₃. Secondly, the evidence of Paiwan /z/ and the reconstruction of the voiceless affricate *C [ɸ] suggests that what we reconstruct as PAN *d₂ [dz] was indeed the voiced counterpart of *C.

Note that the retroflex *d₃ of PAN in Tables 3 and 4 is not identical with the retroflex *d reconstructed by Dempwolff (see Table 2). Dempwolff

reconstructed both a voiceless retroflex *t (Dyen’s *T) and a voiced retroflex *d (Dyen’s *D), neither of which is separately reflected in the Formosan data: they seem to be the result of later developments from PAN *t and *C (Dahl 1981: 22-25) and PAN *d₁, *d₂ and *d₃. It is mildly surprising that a retroflex voiceless stop is not reconstructible in PAN, but the only case of such a segment in Table 3 is Proto-Puyuma *t̪, reflecting PAN *C, and the latter’s reflexes in Saisiyat /s/, Thao /θ/, and Proto-Rukai/Proto-Puyuma *ɸ all imply that PAN *C was dental.

The two contrasting liquid proto-phonemes shown in Tables 3 and 4 as PAN dental/alveolar *L and retroflex *l have been recognised since Dyen (1965a). The difference in point of articulation between these two liquids is clear from Table 3. It is also clear from the table that PAN *l was a lateral, contrasting with PAN *r, a trill at the same dental/alveolar point of articulation. The reflexes of *l in Table 3 imply that it had both lateral [l] and non-lateral [ɿ] allophones. It is unfortunate that the symbol *L has come to be used for the dental/alveolar liquid and *l for the retroflex: this is an artefact of the history of PAN reconstruction.

It is also appropriate to note at this point that the PAN dental/alveolar trill *r contrasted with what was apparently a uvular trill *R. PAN *R has three kinds of reflex throughout much of the Austronesian world: liquid (some variety of r or l), deletion (as in Proto-Paiwan), or a velar (as in Proto-Atayal and many Philippine languages). It is a ready inference from this that *R was a liquid with a weak (easily subject to deletion) dorsal articulation, a description best filled by the uvular trill.

PAN *s and *j (Dempwolff’s *t’ and *g’) have been widely assumed to be palatals, and to be members of a palatal set which also includes, in Dyen’s orthography, *c, *z, and *ñ. However, among the Formosan languages in Table 3, only Proto-Paiwan has phonemically distinct palatal consonants, and these do not reflect PAN palatal proto-phonemes. Hence there is no Formosan evidence for the reconstruction of PAN palatals other than *Z (discussed above),¹⁸ and it appears that the extra-Formosan evidence for these palatal proto-phonemes reflects changes which occurred in PMP and/or after its break-up. Of these putative palatals, *c, *z, and *ñ did not occur syllable-finally, but *s and *j did (Blust 1990b). The segments *c, *z, and *ñ are not distinctively reflected in Formosan languages, whilst Dyen’s *s and *j are better interpreted as dentals [s] and [z]. The following paragraphs show why.

The Proto-Paiwan palatals *c, *j, and *l^y (retained in the Butanglu, Tyuabar and Tyuvuali dialects of Paiwan; Ho 1978) and the dental/alveolars *t and *d represent an interesting reconstructive conundrum. The Proto-Paiwan

palatals correspond with dental/alveolar segments in other Formosan languages and reflect the PAN dental/alveolar proto-phonemes **t*, **d*, and **L*, whilst the Proto-Paiwan dental/alveolars reflect what have traditionally been reconstructed as *palatals* (**s* and **j*). Under this interpretation of PAN phonology, the PAN dental/alveolar and PAN palatal sets have changed places in Proto-Paiwan — a swap for which it is very difficult to find a plausible motivation. If, however, there were no PAN palatals other than **Z*, this problem disappears. Since the Proto-Paiwan palatals reflect the palatalisation (i.e. a backward shift in point of articulation) of PAN dental/alveolars, it is likely that the the Proto-Paiwan dental/aveolars result from a drag chain effect whereby the vacated dental/alveolar slots in the system were filled either by a second backward shift in point of articulation or by a change in manner of articulation, or both. This allows two hypotheses. Either **s* and **j* were articulated further forward than the dental/alveolars, i.e. they were fronted dental or interdental, and/or they were continuants. Apart from Proto-Paiwan, the weight of evidence indicates that their point of articulation was well forward *and* that they were continuants, either [s] and [z] or [θ] and [ð]. Since **s* becomes Atayal, Saisiyat /h/ and Saaroa zero (probably via [h]), and [s] to [h] is a common sound change, [s] is the more natural value to attribute to **s*. Hence we posit the two-stage change below:

- | | |
|---|---|
| 1. PAN * <i>t</i> , * <i>d</i> , * <i>L</i> [l] | > Proto-Paiwan * <i>c</i> , * <i>j</i> , * <i>l̪y</i> |
| 2. PAN * <i>s</i> [s], * <i>j</i> [z] | > Proto-Paiwan * <i>t</i> , * <i>d</i> |

The attribution of a dental value to PAN **j* follows from the Paiwan evidence referred to above, and is supported by its reflection as Thao /ð/ and Saisiyat dental /z/. The velar reflexes in Proto-Atayal and Proto-Rukai are less readily explained, until it is recognised that Proto-Atayal **y* is but one of several possible interpretations of the Atayalic correspondence set, which includes word-medially [r], [s], [w] and [g] and word-finally [y], [s], [?] and [g] (this proto-phoneme, which Li 1981 labels as **g'*, does not occur word-initially). It can be plausibly derived from a lateral fricative, which is perceptually quite similar to and derivable from an interdental fricative.

The proto-phoneme which Dyen (1965a) wrote as **S* subsumed several overlapping correspondence sets, which Dahl (1981: 26-39) has reduced to two, labelled by him **S*₁ and **S*₂. Dahl regards these as two separate phonemes, but this is questionable, since they are reflected identically in Paiwan (as *s*) and in several other Formosan languages. The Formosan sound correspondences for **S*₁ show that it was phonetically something like [s] or [ʃ]. The Formosan reflexes of **S*₂ are either [s], [h] or zero.¹⁹ Since changes

from [ʃ] to [s], from [s] to [h] and from [h] to zero are quite common among the world's languages, it is probable (as Zorc 1982 suggests) that the reflexes of **S*₂ represent a sound change from [s] to [h] in some but not all relevant vocabulary items,²⁰ i.e. that they reflect the same PAN proto-phoneme as **S*₁. Hence they are here both treated as **S*.

If we are correct in attributing the phonetic value [s] to PAN **s*, then **S* must be a sibilant with a different point of articulation. The modern reflexes suggest that it was a lamino-alveolar, [ʃ]. This is certainly possible, but since a distinction between dental/alveolar and retroflex consonants was clearly present in PAN, it is more economic to attribute to it the perceptually similar retroflex value [s].

One small area of controversy remains with regard to the sibilants. Dyen (1971) and Tsuchida (1976) recognise two phonemes **s* and **θ* where I have reconstructed PAN **s*. Tsuchida argues that the two are reflected separately in Maanyan and Malagasy, but Dahl refutes this (1981: 80-82). However, he agrees that the two phonemes must be reconstructed in PAN on the strength of their different reflexes in Kanakanabu, Saaroa and Rukai, with [θ] as the probable articulation of **θ* (1981: 83). I have not accepted this argument, since the three languages all belong to the same putative Formosan subgroup, Rukai-Tsouic.

The hypothesis that PAN did not have a full set of phonemically palatal consonants also receives support from morpheme-structure evidence. Uhlenbeck (1949, 1950) showed that it is a general principle of Javanese morpheme structure that in a morpheme of the canonical shape C₁VC₂V[C], the segments C₁ and C₂ may have the same point of articulation only if they are identical (i.e. *p-b*, *p-m*, *b-m*, *b-p*, *m-p*, *m-b* do not normally occur, but *p-p*, *b-b*, *m-m* do occur). Thus we may say, using an expression from Mester (1988), that the bilabials *p*, *b*, *m* form an "exclusion set". The coronals and velars also form exclusion sets. Adelaar (1983, 1985: 122-125) shows that similar morpheme-structure constraints also applied to Proto-Malayic and probably also to PMP; Clynes (1990) has shown that such constraints occur in Paiwan, and he suggests that they may therefore be reconstructible for PAN. However, if the PMP (and post-PMP) palatals developed from the PAN coronals, we would expect these palatals to form an exclusion set with the coronals, rather than an exclusion set of their own. This hypothesis is confirmed by Mester's (1988: 94, 162-164) statistical analysis of morpheme structure, which shows that in Javanese the coronals and palatals form a single exclusion set.

The reconstruction of the three postvelar phonemes **q*, **h* and **?* may appear odd at first sight. However, such a three-way contrast occurs in

languages scattered in various parts of the world, as the data in Maddieson (1984) show.²¹ Zorc (1982) shows that the three proto-phonemes are reflected by distinct correspondence sets in Taiwan (where Amis and Bunun both reflect them as /q/, /h/ and /θ/), the Philippines and western Indonesia.

The PAN consonant system reconstructed in Table 4 may also appear somewhat eccentric in its overall structure to some readers. However, if a reconstructed phonological system pretends to some sort of realism, we should expect it to resemble the phonology of real languages. Although Dempwolff (1934: 13-23) set out to reconstruct a “symmetrical” system for PMP, and Wolff (1988) attempts something similar for PAN, the consonant systems of real languages are often eccentric in their structures, and the reconstructed PAN system resembles one of these.

The reconstruction of the PAN system of four vowels (*i, *u, *e, *a) and four word-final diphthongs (*-iw, *-uy, *-ay, *-aw) (Table 4) is uncontroversial. Some scholars have suggested that *-ey and *-ew should also be reconstructed (Dyen 1965b, Nothofer 1984), but they are indistinguishable from *-ay and *-aw in Formosan languages, and are therefore not reconstructed here.

Many languages of the Philippines show contrasts between long and short vowels, and Zorc (1978, 1983) has shown that for a number of languages in which the long/short contrast is limited to the vowel of the penultimate syllable this is best explained by the reconstruction of word stress (which, however, he writes as vowel length). Roots either have word stress on the penultimate syllable (reflected in relevant Philippine languages by a long vowel) or on the final syllable (no long vowel occurs in the root). Recent work suggests that Zorc’s system is reconstructible for PAN, since there are relics of it in a few Formosan languages (Ross 1991b).

The canonical shape of PAN roots has long been held to be CVCCVC, where each consonant is optional.²² The majority of reconstructed roots are disyllabic, although trisyllables also occur. Monosyllabic roots (as opposed to grammatical morphemes) occur only in reduplicated form, giving disyllables. Dempwolff (1934: 28) divided roots with a medial -CC- sequence into two categories: those in which the sequence resulted from reduplicating a monosyllable, like *-rg- in PAN **gergér* ‘tremble’, and those in which the sequence consisted of a nasal and a stop at the same point of articulation, like *-nd- in PAN ***pandan* ‘pandanus’ (the double asterisk indicates this is not considered a correct reconstruction). In a number of reconstructions of the latter type, Dempwolff and others since have written the nasal phoneme in brackets because it occurs in some reflexes but not others. Thus we find reconstructions like PAN **ma(n)taq ‘green, unripe’ to account for Paiwan *macak*, Yami *mata*, Isnag *na-māta*, Balinese *matah*, Proto-Oceanic (POC)

**mataq* but Malagasy *manta*, Acehnese *muntah*, Javanese *məntah*; PAN **Cu(m)buq ‘grow, sprout’ to account for Saaroa *ɸuvuʔu* ‘bamboo shoot’, Paiwan *ɸuvuk*, Tagalog *tūbo*, but Malagasy *-tumbu*, Acehnese *tumboh*, Indonesian *bər-tumbuh*, Balinese *tumbuh*, POC **tubuq*; and PAN **qe(n)tut ‘break wind’ to account for Paiwan *kəcuc*, Yami *atot*, Tagalog *utot* but Balanggingi Sama *untut*, Murut *antut*, Indonesian *kəntut*.

Apart from the fact that the theory which allows the insertion of an “optional nasal” is not well motivated, examination of Formosan and other data suggests that the canonical shape of PAN roots was in fact CVCVC, i.e. that the only syllable-final consonants occurred at the end of the root, and that medial consonant sequences in daughter-languages can be accounted for in three basic ways.

The first of these was mentioned above: when monosyllabic roots of the form CVC were reduplicated, they resulted in forms like **gerger*. It is noteworthy, however, that in those Formosan languages which add an extra vowel after an original final consonant (for example, Mantauran Rukai *-kan-ə* and Kanakanavu *-kaən-ə* from PAN **káʔen*), the extra vowel is added after *both* occurrences of the reduplicated monosyllable: Mantauran Rukai *-hər-ə-hər-ə* and Kanakanavu *-kər-ə-kər-ə* from PAN **ger-ger* ‘tremble’. This suggests that in PAN these reduplications were treated phonologically as the repetition of a root, not as a lexical compound.

The second source of medial consonant sequences is PAN trisyllables of the form CVCVC(C) with an unstressed penultimate syllable whose vowel is subsequently lost, giving CVCCV(C). Thus the source of forms like Ilokano *pajdan* ‘pandanus’, Isn *pandān* ‘k. o. screw pine’, Tagalog *pandan* ‘pandanus’, POC **padran* is not a putative PAN ***pandan* ‘pandanus’, but PAN **pajud,al*, as indicated by Formosan reflexes: Budai Rukai *pajodalə* ‘pineapple’, Paiwan *paŋudalʸ* ‘pineapple’. In this example the resulting CC- sequence consists of nasal + stop, but other sequences can also result from this process, which is plentifully exemplified by Blust (1982b).

The third source of medial consonant sequences, referred to in the literature as “nasal infixation”, is less easily explained. Cognate sets like that given with PAN **Cu(m)buq ‘grow, sprout’ require a different explanation from those above. The fact that the “infixed nasal” (Dempwolff called it an “optional nasal”) never occurs in Formosan languages indicates that the PAN form should be reconstructed without it: **Cubuq*. It is noteworthy that items in which the infixed nasal occurs mostly seem to have been PAN verbs, and that it does not occur consistently in Philippine cognate sets, i.e. its presence seems to be the result of borrowing, perhaps from Malay, since the nasal is

prone to occur in languages of western Indonesia. This is a matter which urgently needs research: since the nasal is reflected in some POC items, it was probably (as Dempwolff 1934: 109 notes) the result of a verbal infixation process in a proto-language of quite high order. But since the process does not seem to regularly affect Philippine languages, that proto-language was not PMP. If this is found to be an innovation which defines a large Austronesian subgroup, then it may contribute importantly to our understanding of Austronesian prehistory. Reid (1982) has used this innovation to define a subgroup, and we return to this below.

Considerable space has been devoted here to the reconstruction of PAN phonology, because this is the area of PAN reconstruction in which most work has been done and in which most reconstructive issues have been raised. A much smaller amount of work has been done on the reconstruction of PAN grammar.

Reconstructing Proto-Austronesian grammar

There is an urgent need for more comparative-historical work on Austronesian morphosyntactic and verbal systems. Those of some of the languages of Taiwan, the Philippines and western Indonesia constitute quite unusual structural types among the world's languages. A plausible PAN reconstruction, outlined below, has some of the more unusual features of a number of these languages, and work is needed in order to support or refute this reconstruction. In any case, the PAN system has given rise to a plethora of different systems in modern Austronesian languages, and careful reconstruction of their histories would not only provide copious examples of how grammars change, but languages.

A number of Austronesian languages in Taiwan, the Philippines and parts of western and northern Indonesia retain what was probably the verbal system of PAN: this has come to be known as the system of "topic" and "focus", but will be avoided here. Instead of "topic", we will use the term "pivot", following Foley – van Valin (1984). The following Squliq Atayal examples, from Ferrell (1972), illustrate this system (for a Philippine example, see Tryon,

- a) *k< m>alup saku? bziok ryiax*
<AP>hunt PV:1S boar mountain
'I am hunting boar in the mountains.'

- b) *klup-un maku? ryiax bziok*
hunt-UP GEN:1S mountain boar
'The boar is being hunted by me in the mountains.'
- c) *klup-an maku? bziok ryiax*
hunt-LP GEN:1S boar mountain
'The mountains are having boar hunted in them by me.'
- d) *s-kalup maku? bziok lalau qani*
IPB-hunt GEN:1S boar knife this
'The knife is being used to hunt boar by me.'

The clumsiness of the English sentences below the examples reflects the attempt to capture by quasi-literal translation something of the sense of the Atayal. In this system, each sentence has a pivot (translated as the English subject): the pivot is always definite, i.e. the speaker refers to it in such a way as to indicate that the hearer knows its identity (this is a function of English *the*), and it normally refers to an entity that has been previously mentioned. English, however, has only two subject options — the active *I am hunting* (actor *I* is subject) and passive *The boar is being hunted* (undergoer *the boar* is subject). Atayal, on the other hand, has four pivot options — the actor, the undergoer, the location and the instrument or beneficiary. An affix on the verb indicates what relationship the pivot has to the verb: <*m*> actor pivot (AP), -*un* undergoer pivot (UP), -*an* location pivot (LP) and *s-* instrument or beneficiary pivot (IPB). The noun phrase which is in the pivot relationship to the verb is sentence-final, except when it is a pronoun, as in (a), in which case it occupies the normal pronoun position after the verb, but its form *saku?* indicates that it is the pivot (PV; 1S = first person singular) rather than a non-pivot marked by the genitive (GEN) form *maku?*²³ An important difference from English is that Atayal is a discourse-ergative language: the event line in a narrative is carried by transitive clauses in which the most usual pivot is undergoer, not actor as in English (see McCune 1979, Hopper 1979a, 1979b, 1983, 1988, Wouk 1986, Cooreman – Fox – Givón 1984).

The Atayal system outlined here is very similar to the probable PAN system, which had the four affixes: *<*m*> 'AP', *-*en* 'UP', *-*an* 'LP' and **Si-* 'instrument pivot' (IP) (Starosta – Pawley – Reid 1982).²⁴ The one respect in which the PAN system probably differed from Atayal was that it used not only different pronoun forms but also different noun phrase markers to indicate the relationship of the noun phrase to the verb, thereby allowing greater freedom of phrase order in the sentence. A Paiwan equivalent of (d) above illustrates these markers (Ferrell 1972):

<i>ku</i>	<i>si-kal'yup</i>	<i>a</i>	<i>vuluk</i>	<i>tua</i>	<i>vavuy</i>
GEN:1S	IP-hunt	PV	spear	NPV	pig
'The spear is being used to hunt boar by me.'					

Here Paiwan *a* 'PV' and *tua* 'non-pivot (NPV)' reflect the PAN noun phrase markers **a* 'PV' and **sua* 'NPV' respectively. The instrument pivot marker *si-*, a reflex of PAN **Si-*, indicates that the noun phrase marked by *a* is the pivot of the sentence and is semantically the instrument of the verb.

Starosta – Pawley – Reid (1982) suggest that these affixes were originally nominalisers, which had been or were being reinterpreted as pivot-marking affixes at the time that PAN was breaking up (or they were thus reinterpreted independently in a number of daughter-languages). Under this hypothesis, the phrase *klup-un maku?* meant 'my hunted thing', and the original meaning of the whole sentence was something like 'The boar was my hunted thing of the mountains'. According to Pecoraro's (1979) description of Sediq, a language closely related to Atayal, this ambiguity between nominalisation and verbal pivot marking still exists today. Certainly reflexes of the affixes *-en, *-an, and **Si-* often have meanings derived from their PAN sense, respectively 'something being X-ed', 'place where one X-es' and 'thing one X-es with'. A further affix, PAN *<in>, meant 'something already X-ed', and has assumed the function of a verbal perfective marker in many modern languages.

This set of affixes appears to be historically secondary, and to have been imposed on an earlier, primary system. One reason for thinking that it is secondary is that it consists of a mixture of a prefix, infixes and suffixes, an indicator that a system has been constructed from affixes which originally had other functions. The primary system, which in PAN seems to have been used only for imperative and irrealis verbs, consisted entirely of suffixes: *-a 'AP imperative/irrealis', *-u 'UP imperative', *-aw 'UP irrealis', *-i 'LP imperative', and *-ay 'LP irrealis'. The suffix *-aw, for example, is reflected in Paiwan *ku kal'yup-aw* 'I'll probably hunt' (Ferrell 1982). There has been relatively little discussion of *-a, *-i, and *-ay in the literature and apparently no discussion of *-u and *-aw. Wolff (1973) refers to *-a, *-i, and *-ay, and Starosta – Pawley – Reid (1982) refer to *-a and *-i.

Issues in Austronesian subgrouping

Tryon (this volume) sketches some present views of Austronesian subgrouping. However, the tree diagrams in Figures 1 to 8 are not only subject to the methodological *caveats* discussed earlier: they encapsulate hypotheses which vary widely in quality when they are assessed by the criteria of the comparative method. The method of subgrouping languages on the basis of the innovations which they share relative to a proto language is obviously recursive in its application. Thus Figure 1 shows the primary subgroups of Austronesian languages which developed out of PAN. Figure 2 shows the major subgroups which developed from one daughter-language of PAN, namely PMP. It follows that just as the primary subgroups in Figure 1 need to be determined relative to PAN, so too the highest-order subgroups in Figure 2, namely Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian, need to be determined relative to PMP — and so on down through the tree diagrams in Figures 2 to 8. The purpose of the current section is to give some assessment of the quality of the various hypotheses implicit in those diagrams.

A fairly consistent pattern can be discerned in Figures 1, 2, 6, 7 and 8, which is summed up in Figure 10. At each node is the language of a settled population: the right-hand or right-most branch from each node represents the speech of a segment of that population which migrated from the settled area, so that a new language arose by divergence as a result of that separation. The one case in which a clean separation of the right-hand branch probably did not occur is Proto-Tokelau-Fijian, to which I return below. Together the right-hand nodes represent the main migratory path of Austronesian speakers from Taiwan to Polynesia. What is of note, however, is that most of the left-hand branches do *not* appear to represent a discrete proto language, since they represent the "stay-at-homes". It seems — and we would expect this to be the case — that in most of the cases of separation represented by the nodes in Figure 10, the settled proto language had already diversified into a local linkage before separation occurred. In these cases, the dialects or languages of the stay-at-homes have no exclusively shared ancestor; instead they share only an ancestor at the node above with the language of the departed migrants.

The difficulty of subgrouping the languages of stay-at-homes was examined above with regard to the first Austronesian separation, namely that of Proto-Malayo-Polynesian from the Austronesian linkage on Taiwan. Whilst there is a fair measure of agreement about lower-order subgroups of the stay-at-home languages on Taiwan, there is much less agreement about their higher-order groupings. Scholars agree about the small *Atayalic*,

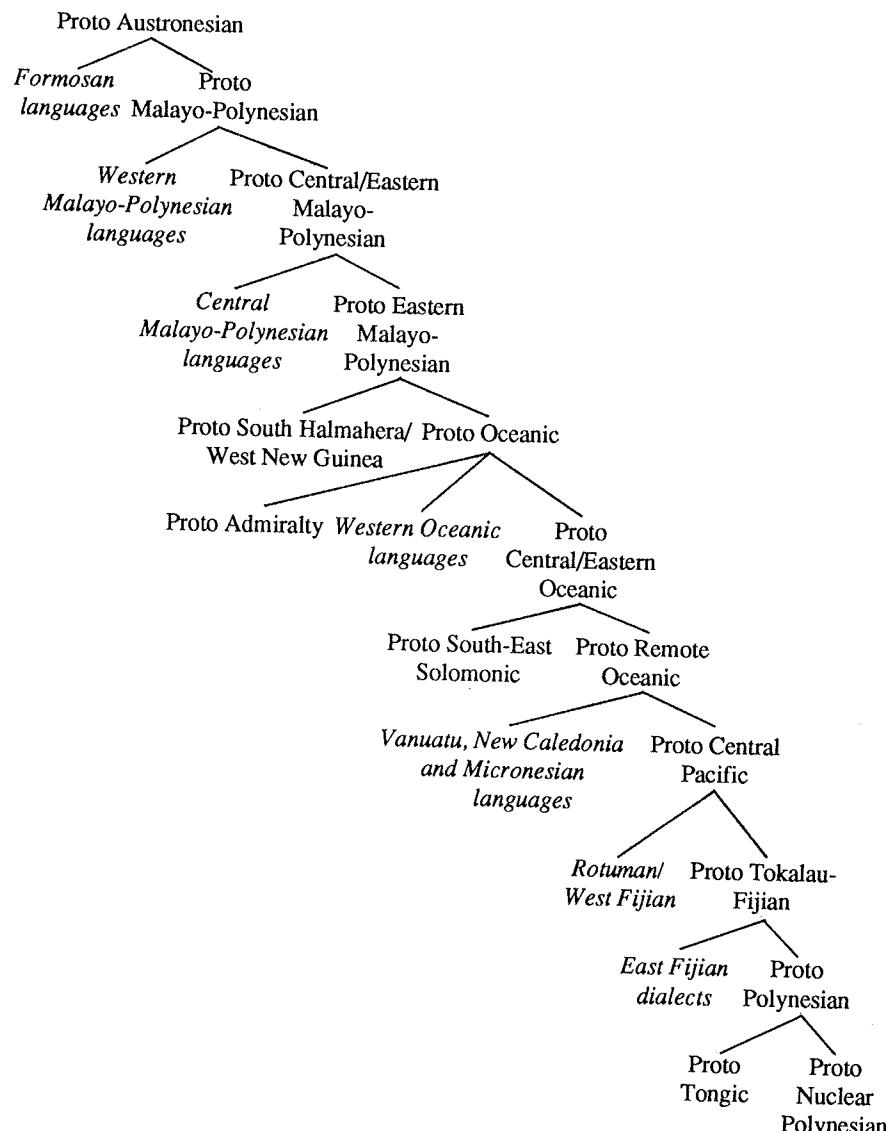


Figure 10. Schematic diagram of the diversification of Austronesian languages

Tsouic, *Rukaic* and *Paiwanic* groups. Detailed comparative work and reconstruction has been carried out for the Atayalic (Li 1980, 1981, 1985), *Tsouic* (Tsuchida 1976) and *Rukaic* (Li 1977) groups, and Tsuchida (1976: 10-12, 216-218) has also shown that the *Rukaic* and *Tsouic* probably group together as *Rukai-Tsouic*. There is general agreement that the *Paiwanic* group consists minimally of *Paiwan*, *Puyuma*, *Amis* and *Bunun*. Beyond this, however, there is little agreement about the subgrouping of Formosan languages, with Tsuchida (1976, repeated in Wurm – Hattori 1983, map 30) and Li (1985) providing quite different genetic trees. The determination of higher-order groups among the Formosan languages is hampered by the fact that many languages have become either extinct or heavily sinicised since the Chinese annexation in the late seventeenth century, with the result that investigators are dependent either on poorly recorded written materials or on data from dying languages. Another difficulty obstructing the subgrouping of Formosan languages is that, because these are stay-at-home languages, it is quite likely that the lower-order subgroups have developed out of a dialect linkage and are *ipso facto* incapable of being subgrouped. Although Tsuchida (1976: 13) points out that the Formosan languages all appear to share certain retentions not shared by Malayo-Polynesian languages, this is not evidence for a Formosan subgroup of Austronesian languages. Rather, the features in which all Malayo-Polynesian languages appear to share innovations are evidence for the unity of the Malayo-Polynesian subgroup.²⁵

Circumstantially, it is almost certain that PAN had diversified into a linkage of dialects and/or languages before speakers of what was to become PMP left Taiwan. It is therefore relevant to ask from which part of the linkage this pre-PMP broke off. In the discussion of Table 1, it was noted that the Formosan languages *Amis* and *Bunun* share with PMP the merger of PAN *C and *t. If one or both of them were found also to share other innovations with PMP, then we would have a probable source for pre-PMP.²⁶ However, research in this direction remains to be done.

The unity of the *Malayo-Polynesian languages* is probably not open to serious question, and Proto-Malayo-Polynesian is readily reconstructible. Blust (1977) argued that Malayo-Polynesian languages share certain innovations in pronoun forms which are not found in Formosan languages. In particular, Malayo-Polynesian languages reflect the “second Austronesian politeness shift”, whereby the morpheme *-mu ‘you (plural)’ came to be used as a polite alternative to *-Su ‘you (singular)’ to mark the possessive. This was apparently followed by the unexplained addition of *i-Su ‘you (singular)’ to *kamu ‘you (plural)’ to form a compound alternant (*kamu-i-Su) to the latter. Another innovation which is reflected right across the

Malayo-Polynesian region is the use of the PMP verbal derivational prefix **paN-*, and **maN-* (< **p<m>aN-* where **<m>* is the actor pivot infix discussed above), to form verbs where the actor is the pivot from verbs where the undergoer is pivot and from other items. The *-N-* represents a nasal segment which often coalesces with the root-initial consonant to form a nasal. In Tagalog, for example, *maN-* combines with the root *sakit* 'be sick' to form the verb *manakit* 'injure', and with *tālo* 'defeated' to form *manālo* 'win (a prize)'. In standard Malay, *məN-* combines with *ikut* 'be followed' to form *məyikut* 'follow'. This innovation is mentioned by Dahl (1976: 127-128) and Reid (1982: 211). Its reflexes are ubiquitous in western Malayo-Polynesian languages. However, they are also preserved in fossilised form in Oceanic languages (Ross 1988: 39-43), indicating that the innovation must have occurred in Proto-Malayo-Polynesian. Another pair of infixes with a similar function is **maR-* and **paR-*, but they occur in this function over a narrower area.

Malayo-Polynesian languages also share in a number of phonological innovations relative to PAN. Ideally, it would be desirable to tabulate evidence concerning the phonetic values of PMP phonemes in much the same way as we did for PAN phonemes in Table 3. However, enough is not yet understood about Malayo-Polynesian subgrouping for this to be feasible, and so Table 5 shows the correspondences between the reconstructed PAN, PMP and POC systems, together with reflexes in four western Malayo-Polynesian languages. The POC system is included because it disambiguates local, western Malayo-Polynesian innovations from PMP.

The conventionally accepted PMP innovations are (Blust 1990b):

- a) PAN *C and *t merged as PMP *t.
- b) PAN *L and *n merged (with some unexplained exceptions)²⁷ as PMP *n.
- c) PAN *S became a glottal spirant of some kind, but did not merge with *h. However, the reconstruction of PAN phonology above rejected some conventionally accepted features on the grounds that they were reflected only in Malayo-Polynesian languages, and not in Formosan, and were therefore PMP (or lower-order) innovations. Hence the list of PMP innovations grows.
- d) PAN *d₁, *d₂ and *d₃ merged as PMP *d.
- e) PMP acquired a palatal nasal *ñ.

Table 5. Correspondences between the PAN, PMP, some western Malayo-Polynesian and the POC phoneme systems

PAN	PMP	Tagalog	Javanese	Toba Batak	Malay	POC
p	p	p	p	p	p	p, b
b	b	b	b,w-b,w-b	b-b-p	b-w-p	p, b
m	m	m	m	m	m	m
w	w	w	w	Ø	h-w-	w
t	t	t	t	t	t	t
C						
s	s	s	s	s	s	s,j
Z	Z	d-l-	d-d-	d-d-	J-J-	s,j
d ₁	d					
d ₂		d-r-d	d,r-d,r-d	d-d-t	d-d-t	r, dr
d ₃						
(z)	(z)	d-r,l-	J-J,r-	J-g-	J-J-	r, dr
(D)	(D)	d-l-d	d,r-d,r-d	d-d-r	d-d-r	r, dr
(T)	(T)	t	ʈ	t	t	t
r	r	l	r	r	r	r, dr
n	n	n	n	n	n	n
L						
l	l	l	l	l	l	l
...	(c)	s	c	s	c	s,j
S	h	h	Ø	Ø	Ø	Ø
j	j	-l-d	r	-g-k	-d-t	c,j
...	ñ	n	ñ	n	ñ	ñ
y	y	y	y	Ø	y	y
k	k	k	k-k-?	h	k-k-?	k, g
g	g	g	g	g-g-k	g-g-?	k, g
ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
q	q	?	h	Ø	h	q
h	-?-H	h	Ø	Ø	Ø	Ø
?						
R	R	g	Ø	r	r	R
i	i	i	i,e	i	i, e, œ	i
u	u	u, o	u, o	u	u, o, œ	u
a	a	a	a	a	a, œ	a
e	e	i, a, u	e	ɔ	œ, a	o
-aw	-aw	-aw	-o	-o	-aw	-o
-ay	-ay	-ay	-e	-e	-ay	-e
-uy	-uy	-oy	-i	-i	-i	-i

The phoneme which Dempwolff symbolised as **d*' and which Dyen writes as **z* is not found in basic vocabulary and has no good Formosan cognates. Its reflexes are limited to western Indonesia, and it is assumed to be a local phenomenon, probably spread through borrowing.

The phoneme which Dempwolff symbolised as **k*' and which Dyen writes as **c* has an even more limited distribution. It is distinguished from **s* only by Javanese, Malay, Acehnese and in loan words in Malagasy. Blust (1988a: 47) argues tentatively that the presence of **c* in six monosyllabic roots (in the sense discussed in the previous section) supports its attribution to PAN, since this is unlikely to be due to borrowing; however, he admits that the argument is weak.

Dempwolff also reconstructed the two retroflex phonemes now written as **T* and **D*. These have clearly separate reflexes (retroflexes) only in Madurese, Javanese and Balinese, and apparent differences in reflexes in other languages are now usually attributed to borrowing from Sanskrit. Blust (1990b) omits them from his paradigms of PAN and PMP reflexes. PMP **ñ* is indisputably reconstructible. It appears to be a PMP innovation, however, since there are no separate reflexes of it in Formosa. A survey of the PAN items containing **ñ* listed by Tsuchida (1976) reveals that there are very few of them and that the Formosan correspondences are more consistently interpreted as reflexes of PAN **n* or **L* (which became PMP **n*) followed by a high vowel.

It was mentioned above that some scholars have reconstructed the PAN diphthongs *-ey and *-ew but that these are not supported by the Formosan evidence. Since they are reflected differently from *-ay and *-aw only in a few languages of western Indonesia (Malay, Sundanese, Madurese and Javanese), they appear to represent a local development, and are therefore also not reconstructible in PMP.

The left-hand branch of Malayo-Polynesian in Figures 2 and 10 is *Western Malayo-Polynesian* (see Maps A and B). As the lack of stratification in Figure 3 implies, we know unsatisfyingly little about the subgrouping of Western Malayo-Polynesian languages, and, as Blust (1985) concedes, there is no evidence that languages assigned to "Western Malayo-Polynesian" form a single Austronesian subgroup (although terms like "Western Austronesian" and "Western Malayo-Polynesian" have often been used in the literature as if these languages did form a genetic unity).

Indeed, Reid (1982: 211-213) has suggested that the *South Mindanao* (his Bilic) subgroup, shown in Figure 3, seems to lack some of the innovations characteristic of Malayo-Polynesian languages and may be descended from "a very early migration south of Formosa by an Austronesian-speaking

people". If this were true, then the South Mindanao group would be co-ordinate with the Atayalic, Rukai-Tsouic and Paiwanic groups of Taiwan. However, the South Mindanao group does reflect the second Austronesian politeness shift, and Blust (1991) points out that some of its atypical features are retentions, not innovations, and are not grounds for exclusion from Malayo-Polynesian. Reid (personal communication) suggests that the idiosyncrasy of the South Mindanao languages in comparison with their neighbours is probably to be attributed to the influence (presumably through language shift) of now defunct Negrito languages spoken in the Philippines before the arrival of Austronesian speakers.²⁸

Students of Philippine languages have traditionally spoken of a "Philippine" subgroup that includes all languages of the Philippine archipelago (except the Sama-Bajaw "sea gypsy" languages), the languages of the Batan Islands between the Philippines and Taiwan, and several groups of languages spoken in the northern arm of Sulawesi, namely Sangiric, Minahasan and Gorontalo-Mongondic. These are often taken to be descended from "Proto-Philippine" (Zorc 1986). However, this grouping has been assumed on typological grounds (as evidenced by the "Proto-Philippine" reconstructions in Wurm – Wilson 1975) rather than justified. Two attempts have been made to justify it on the basis of the comparative method, by Charles (1974), who also includes within the Philippine subgroup the North-East Borneo group, and by Zorc (1986). Reid (1982: 202-203) points out that the innovations which Charles lists as shared by the languages of his Philippine subgroup are based on features of PAN phonology which "do not stand close scrutiny" (and which are not included in the reconstruction of PAN phonology above). Instead, Reid proposes a major Malayo-Polynesian subgroup including all languages in which nasal infixation occurs (see above): he takes this to include perhaps all Malayo-Polynesian languages *except* those of the Northern and Southern Philippine and South Mindanao groups. Since this would leave the Meso-Philippine group and most or all other Malayo-Polynesian languages in the nasal infixation subgroup, but exclude the rest of the Philippines, there could clearly be no Philippines subgroup and no reconstructible "Proto-Philippines".²⁹ This argument is difficult to evaluate, however: as we observed earlier, the nasal infixation process seems to have its centre in western Indonesia and its nature and distribution are not yet known.

Reid's hypothesis is challenged by Zorc (1986), who argues for a discrete Philippine grouping on the basis of a large number of putative lexical innovations shared by languages in the Philippine archipelago. It is also difficult to assess Zorc's position, because of the need to demonstrate that his

lexical innovations are not in fact vocabulary items which have been retained from, say, Proto-Malayo-Polynesian (but lost in extra-Philippine languages) and which have gained currency across the Philippine archipelago as a result of language contact. In this connection it is worth noting Reid's (1982: 212) comment, "As one moves south in the Philippines ... the degree of influence of one or more of the central Philippine languages becomes more and more pervasive, so that it becomes more and more difficult to separate the strata in the languages."

The possible fragmentation of the Philippine languages into genetically distinct Malayo-Polynesian subgroups extends the already lengthy list of groupings which fall under the cover term "Western Malayo-Polynesian". For ease of discussion I will divide the "Western Malayo-Polynesian languages" of Figure 10 into twenty-four groups rather different from those of Figure 3 (the latter are also used in the "Listing of Austronesian languages" in this volume). They are discussed below, together with a statement of the representative languages of each group for which word lists are given in this volume. It should be understood, however, that the grouping together of certain languages does not necessarily entail a commitment to their genetic unity, but is a device for highlighting the "current issues" with which this essay is concerned. The numbers and names in italics refer respectively to map numbers and to language groups in Wurm – Hattori (1981, 1983). The twenty-four groups are:

- 1) Batanic (30, 32: *Batanic or Ivatan group*)
- 2) Northern Philippines (32: *North Philippine group*)
- 3) Meso-Philippines (32-33)
- 4) Southern Philippines (33-34)
- 5) South Mindanao (34)
- 6) Chamorro (Mariana Islands) and Palauan (Belau) (18)
- 7) Sangiric (Talaud and Sangir Islands and north-eastern Sulawesi) (34, 43)
- 8) Minahasan (north-eastern Sulawesi) (43)
- 9) Gorontalo-Mongondic (northern Sulawesi) (43)
- 10) Sama-Bajaw (Sulu Archipelago and other scattered locations in the Philippines, East Malaysia and Indonesia) (33, 34, 40, 43-45)
- 11) Malayo-Chamic (35-39, 41-42: *Acehnese, Chamic, Malayan, Sundanese, Madurese*)
- 12) Moken and Moklen (islands off the west coast of Thailand and Myanmar (formerly Burma)) (36)
- 13) North-West Borneo (41, 42: *North-East Borneo, Rejang-Baram, Kayan-Kenyah groups*)

- 14) Land Dayak (inland south-west Borneo) (41)
- 15) East Barito (south-east Borneo and Madagascar) (42)
- 16) Barito-Mahakam (south-east Borneo) (42)
- 17) West Barito (southern Borneo) (42)
- 18) Lampung (south-east Sumatra) (38)
- 19) North-west Sumatra/Barrier Islands (38: *Gayo, Batak, Nias, Mentawai, Enggano*)
- 20) Java-Bali-Sasak (39-40)
- 21) Central Sulawesi (43-44)
- 22) South Sulawesi (43-44)
- 23) Muna-Buton (islands off south-east Sulawesi) (44)
- 24) Tamanic (central Borneo) (41: *Mbaloh*)

The first five groups were discussed above. Zorc (1978) combines the Batanic subgroup (which includes Yami) and the Northern Philippine subgroup (including Isneg and Limos Kalinga) shown by McFarland in Wurm – Hattori (1983). The *Meso-Philippines* group is equivalent to Zorc's (1978) Central and Meso-Philippine groups and includes Tagalog, Aklanon Bisayan, Palawan and Molbog. The *Southern Philippines* group includes Kagayanan Manobo, the *South Mindanao* group Sarangani Blaan. The most thorough comparative work on a Philippines group is Zorc's (1977) study of the (Meso-Philippine) Bisayan dialects. More piecemeal comparative work has also been published for the Meso-Philippine East Mindanao subgroup (Gallman 1979, Pallesen 1985: 125-156, 289-297) and for subgroups within the Southern Philippine group (Elkins 1973-74, 1983-84, Allison 1979, Harmon 1979, Fleischmann 1981).

The languages of Groups 6-10 have generally been associated by scholars with languages of the Philippines (hence Gorontalo, group 9, is included with the Southern Philippines languages in Figure 3). However, this association is based on the shared retention of PMP morphosyntactic features rather than on shared innovation, and these are not grounds for subgrouping.

Chamorro and *Palauan* are combined in Group 6 because both belong geographically to Micronesia, and are situated to the east of the Philippines. However, they share no known innovations, and little can be said about their genealogy other than that they are clearly Malayo-Polynesian but do not belong to the Central/Eastern Malayo-Polynesian group (i.e. they are "Western Malayo-Polynesian"). Starosta and Pagotto (1991) have suggested on syntactic grounds that Chamorro's closest relatives could be either in Taiwan or the Philippines. However, both Chamorro and Palauan reflect the second politeness shift and verbal derivation with *mag-: they are therefore Malayo-Polynesian, and their closest relatives cannot be on Taiwan. The

difficulty in assigning these two languages to a subgroup or subgroups of Malayo-Polynesian lies in our ignorance of many western Malayo-Polynesian languages: good descriptions of both Chamorro (Topping 1973, 1975) and Palauan (Josephs 1975, McManus 1977, Hagège 1986) are available.

The languages of Groups 7, 8 and 9 are all located in the region which includes northern Sulawesi and the islands to its north towards Mindanao. The proto-languages of the three groups have been reconstructed and their implications for subgrouping examined, for *Sangiric* by Sneddon (1982), for *Minahasan* by Sneddon (1978) and for *Gorontalo-Mongondic* by Usup (1986) and Sneddon – Usup (1986). Group 9 is represented here by Gorontalo. Sneddon (1989) concludes that no innovations can be identified which are shared by any of the groups with each other or with any of Groups 1-4.

Sama Bajaw (Group 10) is a collection of dialects spoken by “sea nomads” who live in a symbiotic relationship with various more powerful land-based groups scattered across the central and southern Philippines (and especially locations in the Sulu Archipelago, from which the Balangingi Sama word list is drawn), Sabah, Sulawesi and the Moluccas (Pallesen 1985, Verheijen 1986: 29-30). Proto-Sama-Bajaw has been reconstructed by Pallesen, who believes that it has its affiliations in Indonesia (1985: 245-246), but work to show where has yet to be done.

Group 11, the *Malayo-Chamic* group, covers substantial areas of western Indonesia and West Malaysia. It is represented by word lists from Aceh, Minangkabau, Indonesian, Sundanese, and Madurese (the geographic “Sundic” group of Figure 3 includes languages from Groups 19 and 20, as well as Group 11). The Malayo-Chamic group is proposed by Blust (1981a, 1985), but needs further work for its definition in terms of shared innovations. Within the group, two major subgroups can be distinguished, Malayic and Aceh-Chamic.

The internal subgrouping of *Malayic* is not yet fully understood, but it includes the subgroup which Wurm (in Wurm – Hattori 1983) has labelled “Malayan” (centred on the Malay Peninsula, east Sumatra, and southern Borneo), and possibly Sundanese (west Java) and Madurese (Madura Island, off east Java). However, the Malayic subgroup is not without controversy. Nothofer (1988) proposes on the basis of lexical innovations that Tamanic (his “Maloh”, our Group 24) also belongs to it, but that it excludes Madurese.³⁰

Adelaar (1985) provides a well founded reconstruction of the proto-language of “Malayan”, somewhat confusingly named “Proto-Malayic”. He

argues that this proto-language was probably spoken somewhere in western Borneo, in the region where today’s Malayic Dayak languages are spoken (Adelaar 1988, forthcoming). According to Blust (1988b), by about two thousand years ago “Malayan” speakers had also established themselves on both coasts of the Strait of Malacca (on the Malay Peninsula and in the south-eastern half of Sumatra), and from about a thousand years ago settled various enclaves in central and eastern Indonesia, giving rise to such varieties as Ambonese Malay, Bacan Malay and Menado Malay.

The integrity of the *Aceh-Chamic* group has been demonstrated by Cowan (1948, 1974, 1988) and Durie (1989). The Chamic languages are spoken in inland relic areas in southern Vietnam and Kampuchea (and one language on the Chinese island of Hainan), and are surrounded by languages of the Austro-Asiatic Mon-Khmer family which have apparently reduced what was once a much larger Austronesian speaking territory. Aceh of north-west Sumatra represents a colony established from this mainland region and, like the Chamic languages, has undergone changes resulting from contact with Mon-Khmer languages.

Group 12 consists of the dialects of the *Moken* and *Moklen* “sea gypsies” along the coast of the Andaman Sea. Reid (pers. comm.) suggests that this language is probably a member of the Aceh-Chamic group, as it also has Mon-Khmer phonology: it is basically monosyllabic, with a “pre-syllable”, whereby full vowel contrasts occur only in the full syllable, not in the pre-syllable. Moken/Moklen has often been confused with Urak Lawoi’, the language of the sea gypsies of the islands to their south (Hogan 1988: 1-2). The latter, however, is a “Malayan” dialect belonging to Group 11.

The geographic “Borneo” group in Figure 3 includes the area of Groups 13-17 and 24 in the present classification.

The unity of Group 13, *North-West Borneo*, all of whose members are situated on the island of Borneo, is argued on the basis of shared innovations by Blust (1974b, 1985, pers. comm.). However, Adelaar (pers. comm.) suggests that Rejang-Baram may belong with Group 14. Group 13 is represented here by Timugon Murut.

The genealogy of the *Land Dayak* languages of south-west Borneo (Group 14) remains unknown, but Adelaar (1991) gives evidence that, like the Aceh-Chamic languages and Moken/Moklen, these languages show signs of contact with Mon-Khmer languages: they may be the result of language shift by a group of the Mon-Khmer speakers related to those of the Malay Peninsula.

Groups 15, 16 and 17 were combined by Hudson (1967) into the *Barito* group. However, in his 1978 paper they are split into three without

explanation, and there are no known grounds for combining them. Dahl (1977) reviews the phonological evidence for Hudson's groupings and largely agrees with them, but offers no qualitative evidence for the larger Barito grouping. East Barito (Group 15) is well known for the fact that one of its members is Malagasy, the language (or group of languages) spoken on Madagascar, as Dahl (1951) has shown. Until recently, it has been accepted that Barito speakers arrived on Madagascar by the fourth or fifth century A.D., but recent work by Adelaar (1989) on Malay loan words in Malagasy indicates that the migration must have occurred between the fifth and the thirteenth centuries, with the seventh century A.D. as the most probable period. No detail is available with regard to the genetic affiliations of Groups 15-17, nor of Group 18.

Group 18 contains only *Lampung*, of extreme south-east Sumatra. Although it has been suggested in the past that it belongs to the Malayic group, current opinion regards it as not yet classified (Blust, pers. comm., Nothofer 1988). Comparative dialect material and a short grammar of one dialect have been published by Walker (1975, 1976).

Although the languages of *north-west Sumatra* and *the Barrier Islands* off Sumatra's south-west coast (Group 19) appear superficially quite diverse, there is comparative evidence that they form a single genetic group. Adelaar (1981) demonstrated the unity of the Batak languages and reconstructed Proto-Batak, and Nothofer (1986) has shown that the members of the group share a number of innovations in common. The best described language of the group is Toba Batak, for which a word list is provided here.

Group 20, the *Java-Bali-Sasak* group, was proposed by Esser (1938), but he provides no supporting evidence for it. The Bali-Sasak connection is well established, but its relationship to Javanese requires formal demonstration. It is possible that similarities between Balinese and Javanese owe more to intense contact before the sixteenth century A.D. islamification of Java than to shared inheritance. It has also often been assumed that Javanese is quite closely related to Malayo-Chamic, but again the similarities may well be largely due to contact. Hopefully, the inclusion here of word lists for Javanese, Balinese and Sasak will assist scholars working in this area. Linguistic study of Old Javanese and Old Balinese texts is also needed in order to reconstruct their linguistic history and the effects of contact, although it is probable that the texts are linguistically conservative in relation to the spoken languages contemporaneous with them.

Groups 21, 22 and 23 are all spoken on or near the island of Sulawesi, and are shown as a "Sulawesi" group in Figure 3. The *South Sulawesi* group is well established, as a result of Mills' (1975) work, but the integrity of the

Central Sulawesi and *Muna-Buton* groups is uncertain (C. Grimes, pers. comm., van den Berg 1991). The term "Muna-Buton group" is due to Esser (1938), who includes in it the Wolio language, for which a list is included in the present work. Relationships among the languages attributed to the three groups are unclear, although impressionistically those of the South and Central Sulawesi groups seem to be connected. The inclusion of two Central Sulawesi (Da'a, Uma) and two South Sulawesi (Bugis, Konjo) lists in the present work will improve this situation.

Interestingly, Adelaar (1991, forthcoming) has recently presented a set of innovations shared by the *Tamanic* languages of Central Borneo (Group 24) and the languages of South Sulawesi (Group 22), and proposes a genetic relationship between them.

Blust (1985 and pers. comm.) suggests that Malayo-Chamic, the three Barito groups, and Java-Bali-Sasak are all descended from a language which we may call Proto-South-East Borneo. This language separated into Proto-Barito, ancestral to the Barito groups, and a proto-language ancestral to Malayo-Chamic and Java-Bali-Sasak.

The broadest-brush approach to the subgrouping of "Western Malayo-Polynesian" languages is that taken by Nothofer (1991), who suggests that much of the western Malayo-Polynesian region was once occupied by speakers of languages belonging to a group which he dubs "Palaeo-Hesperonesian". At a later date, much of this area was then occupied by speakers of "Hesperonesian" languages, who became culturally dominant, in western Indonesia and West Malaysia, displacing Palaeo-Hesperonesian languages, which today survive around the periphery of the Western Malayo-Polynesian region. If I interpret Nothofer correctly, the five Philippine groups (1-5), the six Sulawesi groups (7-9, 21-23), the north-west Sumatra/Barrier Islands group, and probably the North-West Borneo group would all be assigned by him to the Palaeo-Hesperonesian grouping, whilst Malayo-Chamic, the three Barito groups, and Java-Bali-Sasak (i.e. the higher-order grouping also proposed by Blust) form the Hesperonesian grouping.³¹ Much of the evidence which Nothofer uses is lexical, and subject to the same difficulties as Zorc's use of lexical evidence in the Philippines. Nonetheless, his proposal would do much to make sense of the chaos in western Malayo-Polynesian classification, and deserves careful investigation by the comparative method.

Reid (pers. comm.) points out that the Meso-Philippine languages and those which Nothofer groups as "Hesperonesian" exclusively share an innovation whereby the PAN ligature *(n)a (Starosta – Pawley – Reid 1982) has become -ŋ, reflected for example in Tagalog *ay* 'pivot common noun

phrase marker' (< **a-ŋ* noun phrase marker + ligature), Indonesian *yaj* 'relative clause marker', Aceh *ñag* 'relative clause marker', and the Old Javanese ligature *-ŋ*. This suggests, among other things, a possible modification to Nothofer's scenario, in that it would group the Meso-Philippine languages with his Hesperonesian. However, the split in this hypothesised group would have had to occur before PAN **R* became Proto-Meso-Philippine **g*.

The confusion which shrouds the subgrouping of western Malayo-Polynesian languages makes it difficult to reconstruct much of what happened immediately after speakers of an early Austronesian communalect migrated southward from Taiwan to the Batan Islands or to Luzon and their speech became Proto-Malayo-Polynesian. Proto-Malayo-Polynesian speakers probably spread through the Philippines, where their speech gradually diversified into a number of languages. However, the languages of the Philippines today are less diverse than we would expect, and Blust (1985) suggests that at various times in the past five thousand years what he calls "linguistic levelling" has occurred, i.e. one prehistoric language has expanded at the expense of its neighbours and/or massive contact-induced change has occurred, reducing linguistic diversity. It is perhaps this situation which has given rise to controversy over Philippine subgrouping, as scholars have attempted to separate archaic features and shared innovations from borrowings and other contact-induced changes.

There are three general routes by which speakers of Malayo-Polynesian communalects could have entered the Indo-Malaysian archipelago from the Philippines, carrying with them their rice-growing culture: through north-east Borneo, through northern Sulawesi, and through Halmahera in the northern Moluccas.

The first of these routes was certainly used, and a Malayo-Polynesian language was spoken in north-west Borneo, perhaps by 2500 B.C. when rice was already growing in the neighbourhood of the Gua Sireh archaeological site in Sarawak (Bellwood pers. comm.). The North-West Borneo group may well consist of descendants of those first Austronesian settlers, who have remained more or less *in situ*, or it may be the result of a later incursion into Borneo. Some Malayo-Polynesian speakers probably interacted with Mon-Khmer speaking hunter-gatherers, to judge from the evidence of Land Dayak. More importantly, however, they established a strong presence in the west of Borneo, where the Barito and Malayic Dayak languages are now located, and on both Blust's and Nothofer's hypotheses this area was the centre from which the languages ancestral to the Malayo-Chamic and Java-

Bali-Sasak groups originated. However, this is a speculative reconstruction which requires a great deal of both linguistic and archaeological support.

The second route was evidently also used and Malayo-Polynesian communalects entered Indonesia through northern Sulawesi. It is unlikely that all the groups in Sulawesi are descended from a single Malayo-Polynesian migration. The three northern groups (7-9) in particular may be descended from more recent incursions.

Clearly these speculations would require some adjustment to fit Nothofer's Palaeo-Hesperonesian hypothesis and/or Reid's Meso-Philippine/South-East Borneo connection. All these proposals require detailed data collection and careful comparative work for their confirmation or refutation.

It is unclear whether the third route into Indonesia, through Halmahera, was used at an early date: speakers of a Malayo-Polynesian communalect apparently entered the Moluccas between 3000 and 2500 B.C., and their speech became Proto-Central/Eastern Malayo-Polynesian (PCEMP). But whether they came from the Philippines directly to Halmahera or, as seems geographically more likely, through Sulawesi is not yet known. The speakers of PCEMP dispersed quite rapidly, one group settling probably on Halmahera, where their language evolved into Proto-Eastern-Malayo-Polynesian (PEMP), and another group occupying a location further south, where their language became Proto-Central-Malayo-Polynesian (PCMP). From here, sometime before 2000 B.C., Central Malayo-Polynesian speakers spread across the south Moluccas and the eastern part of Nusa Tenggara (the lesser Sunda Islands; see Map B).

Of all the right-hand branches in Figure 10, the Central/Eastern Malayo-Polynesian (CEMP) is the least well supported by the comparative method. The grouping was proposed by Blust (1974a) and evidence put forward in its support by Blust (1982a, 1984, 1990a).

In relation to the reconstruction of PMP phonology outlined above, PCEMP underwent no innovations in its phoneme system. However, all CEMP languages reflect a reduction of heterorganic consonant clusters arising through reduplication, except where this resulted in a nasal + stop sequence. For example, PMP **bukbuk* 'wood weevil' became PCEMP **bubuk*, but PMP **demdém* 'dark' became PCEMP **dendem*. There are also a number of irregular phonological changes which occurred in single words, although as Blust (1990a) admits, some of these may be the result of independent parallel innovation (e.g. many languages reflect **sei* 'who?' rather than **sai*). The irregular change from PMP **maRi* 'come' to PCEMP **mai*, on the other hand, is significant.

Blust (1990a) also provides a number of PCEMP lexical and semantic innovations. The most telling of these were published in Blust (1982a), where he argues that as Malayo-Polynesian speakers crossed the Wallace Line, which divides the region occupied by placental mammals from the region inhabited by marsupial mammals, they required names for the new animals they encountered. Since languages in both the Central Malayo-Polynesian and Eastern Malayo-Polynesian groups (including some in the Oceanic subgroup) reflect the etyma **kandoRa* ‘cuscus’ and **mansar* ‘bandicoot’, he argues that there must have been a PCEMP language in which these innovations occurred.

In Figures 2 and 10 the left-hand branch of CEMP is Central Malayo-Polynesian. As with the previous left-hand branch, Western Malayo-Polynesian, there is also little evidence that the Central Malayo-Polynesian languages form a closed subgroup. It is probable that PCEMP had diversified into a local linkage before PEMP separated from it, leaving the “stay-at-home” Central Malayo-Polynesian languages with no exclusively shared ancestor. Blust (1990a) describes a number of phonological and lexical innovations which seem to have originated in different parts of the early Central Malayo-Polynesian linkage and to have spread through part, but not all, of the linkage. The result is patchily distributed innovations and differences in rule ordering among languages. Thus many Central Malayo-Polynesian languages show what Blust calls “glide truncation”: the final glide of inherited **-ay*, **-aw*, or **-uy* is deleted, leaving *-a* or *-u* (for example, PMP **matáy* ‘die’ > Buru *mata*, PMP **pánaw* ‘walk’ > Buru *pana*, PMP **bábuy* ‘pig’ > Buru *fafu*).³² But this innovation does not occur in all Central Malayo-Polynesian languages, nor in all eligible items in languages where it does occur. One possible innovation that Central Malayo-Polynesian languages have in common is their sharing of the pronominal subject proclitic forms **ku-* ‘1S’, **mu-* ‘2S’, **na-* ‘3S’, **ma-* ‘1EP’, **ta-* ‘1IP’, **mi-* ‘2P’, and **da-* ‘3P’ (Blust 1990a). However, the fact that sets of very similar proclitic forms occur in western Malayo-Polynesian and South Halmahera languages casts doubt on this as an innovation.

Recognised groupings of Central Malayo-Polynesian languages are listed here under the same conventions as were used above for Western Malayo-Polynesian (the first five are as shown in Figure 3).

- 1) Bima-Sumba (eastern part of Sumbawa, Sumba, Flores) (40)
- 2) Timor (40: *Timor and Waima'a groups*)
- 3) South-East Maluku (Tanimbar and Kai island groups) (4, 40)
- 4) Aru (4)
- 5) Central Maluku (Seram, Buru, and their offshore islands) (45)

- 6) North Bomberai (south coast of MacCluer Gulf, Irian Jaya) (2)
- 7) Koiwai (south coast of Bird's Neck, Irian Jaya) (2: *Kaiwai*)

These groups form the basis of Figure 4. I am indebted to Charles Grimes, who provided most of this classification and is the source of much of the commentary below.

Group 1, the *Bima-Sumba group* is represented in this work by word lists from Manggarai and Ngada. The Bima language on Sumbawa Island is at the interface between the Bima-Sumba group and the Western Malayo-Polynesian Sumbawa language (closely related to Sasak and Balinese). Although the Bima-Sumba group has long been recognised (by Esser 1938, who followed Jonker's work thirty years earlier), no work appears to have been done to establish its validity in terms of shared innovations.

The *Timor group* (Group 2) is represented in this work by Sikka and Roti. It matches Fox and Wurm's “Timor Area group” in Wurm – Hattori (1983) with the addition of the Waima'a languages, which appear superficially rather different from other Timor languages because of contact-induced change associated with their locations, surrounded by Papuan languages. Again, there is no published work presenting shared innovations of this group.

The *South-East Maluku group* (Group 3) is as recognised by Wurm – Hattori (1983) and by Collins (1982) but with the exclusion of the languages of the *Aru Islands*, treated here as Group 4 and represented in this work by a Dobel word list. Collins places the languages together because of their shared merger of PMP **R*, **r*, and **j*, but recognises that the Aru languages are otherwise quite different from those of Tanimbar and Kai. However, this innovation occurs elsewhere in the region and is therefore not exclusively shared by these languages.

The *Central Maluku group* (Group 5) has long been recognised, as Collins (1983) shows in his survey of Central Maluku studies. Collins (1981) shows that, in terms of the reconstruction of PMP adopted here, Proto-Central Maluku merged PMP **mb* and **mp* as Proto-Central Maluku **mb*, and PMP **nd* and **nt* as Proto-Central Maluku **nd*. PMP **j* became Proto-Central Maluku **-t* word-finally. He also (1983) identifies some morphological innovations. However, in both works he remarks that this is not a large collection of innovations, and that further research is needed to determine whether these innovations are exclusively shared by the members of Group 5 or whether they define some higher-order group. In the latter case, they would cease to be grounds for recognising the Central Maluku group, although most or all of the groups which Collins has defined *within* Central Maluku would continue to stand. Collins (1982) is hesitant to group the

Teor-Kur languages (to north-west of Kai Islands) with Group 5 because they fail to undergo some of the innovations of their neighbours. However, they reflect the innovations which define Central Maluku. It is possible that Groups 3 and 5 will eventually prove to be descended from an ancestral chain.

Groups 6 and 7 are assigned by Wurm – Hattori (1981, 1983) and in Figure 5 to the South Halmahera/West New Guinea group of Eastern Malayo-Polynesian languages. However, Blust (1990a) shows that they reflect Central, rather than Eastern, Malayo-Polynesian innovations. Group 6, which he labels “*North Bomberai*” includes the Sekar, Onin, and Uruangnirin languages, and possibly Arguni, as well perhaps as other nearby Austronesian languages for which no data are available. Blust also argues for a subgrouping relationship between the North Bomberai group and Yamdena, here assigned to the South-East Maluku group, and suggests that Groups 3, 4, and 6 (and perhaps other languages too) will eventually prove to form a higher-order grouping.

Group 7 consists of a single language, *Koiwai*, spoken by an Islamicised group in the south coast of Irian Jaya.

The comparative method has as yet been little applied to questions of subgrouping within the Central Malayo-Polynesian branch. This is an area in which there is enormous scope for research.

The right-hand branch of the PCEMP node in Figure 10 is occupied by PEMP. It is again one of Blust’s papers (1978) which is the source of this subgrouping hypothesis. Evidence for the Eastern Malayo-Polynesian grouping consists essentially of putative lexical innovations (Blust gives 56), convincing phonological innovations shared in common by the two member groups, South Halmahera/West New Guinea (SHWNG) and Oceanic (see Maps A and B).

PEMP in its turn gave rise to two languages, Proto-SHWNG and Proto-Oceanic (POC).

Proto-SHWNG is one of the few cases in Figure 10 where we find at the left-hand branch a single proto-language, as opposed to a group of “stay-at-homes” who have no exclusively shared ancestor. Because the unity of Proto-SHWNG is remarkable, it is worth summarising its phonological innovations here, relative to the PMP phonology reconstructed above. They are:³³

- a) PMP *p became Proto-SHWNG *f
- b) PMP penultimate *e became Proto-SHWNG *o

- c) PMP initial *a- became Proto-SHWNG *ya-
- d) PMP *t became Proto-SHWNG *s before *i
- e) PMP *-j- merged with *s as Proto-SHWNG *-s-
- f) PMP *k, *q, *h, *H and *? were lost
- g) PMP *d, *Z, *l and *r merged as Proto-SHWNG *l
- h) PMP *n and *ñ merged as Proto-SHWNG *n
- i) PMP final syllable *e merged with final syllable *a as Proto-SHWNG *a

Although some of these innovations (a, c, d, f, h) occur with sufficient frequency in Austronesian languages for us to infer that they could have occurred independently after the break-up of Proto-SHWNG, others (b, g, i) are quite striking and their occurrence in combination proves the integrity of Proto-SHWNG as a language. Blust (1978) also demonstrates the integrity of each of the two SHWNG subgroups, and my unpublished research bears this out: Proto-SHWNG split into Proto-South-Halmahera and Proto-West-New-Guinea. The South Halmahera group is represented in this work by Sawai and Irarutu.

The fact that PEMP split into two discrete proto-languages and did not give rise to a collection of “stay-at-homes” casts some light on why it is defined by no exclusively shared innovations outside the lexicon. It indicates that PEMP did not diversify into a linkage of dialects before it split, but either spent a relatively short period as a discrete language or spent its time as a discrete language in a small and remote location where the possibility of dialect differentiation was very limited. Positing the smallest number of migratory moves, this location was either Halmahera or the Cenderawasih Bay area, to the east of the Bird’s Head of Irian Jaya — more probably the former, on the principle that the location of the more conservative members of a language group is likely to be its homeland (Ross 1991c). Either Halmahera or the Cenderawasih Bay qualifies as an isolated location. It was probably from here sometime around 2000 B.C. that speakers of the immediate ancestor of POC made their way via the offshore islands of the north coast of Irian Jaya and mainland Papua New Guinea to a new homeland in the Bismarck Archipelago. The people they left behind them spoke a virtually undifferentiated form of PEMP, which underwent a series of phonological innovations to become Proto-SHWNG before some of its speakers moved from Halmahera to the coast and islands of Cenderawasih Bay or vice versa.

In his attempt to delineate the Irian Jaya boundary between Central and Eastern Malayo-Polynesian languages, Blust (1990a) has questioned whether Irarutu, one of the languages for which a word list appears in this work and

which has previously been attributed to the SHWNG group, belongs to SHWNG or to Central Malayo-Polynesian. Since the latter is ill-defined, realistically we can only ask: is Irarutu a SHWNG language? The answer appears to be that it is. Because Irarutu has undergone extensive lexical replacement, probably through contact with a neighbouring Papuan language or languages, it is often difficult to find relevant examples. However, there is evidence of the SHWNG innovations (a), (b), (f) and (g) as follows:

- a) PMP *penúq 'full' > Irarutu *nə-phonə*
PMP *páqa 'thigh' > Irarutu *fa* 'leg'
- b) PMP *telú 'three' > Irarutu *torə*
PMP *degeR 'hear' > Irarutu *nafa-roŋ-kərə* (and *penuq above)
- f) PMP *káyuh 'tree' > Irarutu *ɛ*
PMP *kútuh 'louse' > Irarutu *utə*
PMP *panákaw 'steal' > Irarutu *mbana*
PMP *iakú 'I' > Irarutu *ja*
PMP *qatéluR 'egg' > Irarutu *təru* (and *páqa above)
- g) PMP *dáhun 'leaf' > Irarutu *rɔ*
PMP *sidá 'they' > Irarutu *irə* (and *degeR above)
PMP *Zálan 'path' > Irarutu *randəni*
PMP *lágit 'sky' > Irarutu *ragətə*
PEMP *bitəl 'hungry' > Irarutu *mbitərə* (and *qatéluR above)

Although these groupings are well established on the basis of the available data, there remain languages in Halmahera for which the only published information is that contained in the Holle lists (Stokhof (ed.) with Saleh-Bronckhorst 1980), lists collected by Dutch amateurs (mainly colonial officials), and languages around the Bird's Head and Cenderawasih Bay area of Irian Jaya for which we have no data at all.

Although the Oceanic subgroup of Austronesian languages covers a larger area than any of the other major groups, its integrity is quite clear and its internal subgrouping reasonably well understood. Table 6 sets out the correspondences between the reconstructed PMP and POC phoneme systems. The POC symbols have their expected values, except that voiced stops were prenasalised. The development of the POC consonant system is analysed in some detail in Ross (1988, chapter 3). The table also shows the equivalences between the POC orthography used here (from Ross 1988) and that used by Grace (1969) and other scholars.³⁴

Table 6. Correspondences between the reconstructed phoneme systems of Proto-Malayo-Polynesian and Proto-Oceanic

PMP	p, b	t	d, r	s, Z	j	k, g				
POC	oral p nasal b	p ^w b ^w	t d	r dr	s j	c j g				
Grace	oral p nasal mp	... ŋp	t nt	d nd	s nj	j nj ŋk				
PMP	m	n	ñ	ŋ	w	y	l	q	h, H, ?	R
POC	m	m ^w n	ñ	ŋ	w	y	l	q	∅	R
Grace	m	ŋm n	ñ	ŋ	w	y	l	q	∅	R
PMP	i, -uy	e, -aw	-ay	a	u					
POC	i	o	e	a	u					

The innovations which occurred in POC were a set of mergers and splits. The PMP consonant pairs *p/*b, *d/*r, *s/*Z and *k/*g each merged in POC. However, where a PMP obstruent was preceded by a nasal in the language immediately ancestral to POC, a unitary prenasalised voiced consonant resulted. Thus, for example, PMP *k and *g merged as POC *k [k], but earlier *ŋk and *ŋg merged as POC *g [g]; POC *k is known as an oral grade obstruent, whilst *g is its nasal grade equivalent. I write "earlier" rather than "PMP" because PMP probably had no word-initial nasal + stop sequences, whereas the immediate ancestor of POC certainly did. For example, PMP *kúden 'cooking pot' became POC *kuron, but PMP *kabut 'mist' became POC *gabu(t), evidently via a Pre-POC form *ŋkabu(t). In the case of the PMP bilabials an additional complication intervened, as they split into plain and labialised bilabials. PMP *e became POC *o, and the PMP word-final diphthongs merged with plain vowels.

It is very probable that there were morphosyntactic innovations in POC. However, although we can claim to understand POC morphosyntax in some measure (Pawley 1973, Pawley – Reid 1980, Lichtenberk 1985a, 1985b,

Crowley 1985, Ross 1988, chapter 4), we do not know enough about PEMP grammar to reconstruct the innovations of POC with reasonable certainty. There were also numerous lexical innovations in POC, but no list of these has ever been compiled. For example, PMP **limaw* ‘citrus fruit’ was replaced by POC **molis*.

It seems unlikely (but not impossible) that the immediate ancestors of POC speakers should have migrated from Halmahera or the Bird’s Head directly to somewhere in the Bismarck Archipelago, and there is no need to posit a single direct migration. There is linguistic evidence that Pre-POC speakers were in contact with Papuan speakers on the coast and offshore islands of the Madang area of Papua New Guinea’s north coast, perhaps *en route* to the Bismarcks (Ross 1988: 21), and the contiguity of several Oceanic subgroups in the Bismarck Archipelago makes it the most likely location for POC. I have argued elsewhere that the north of New Britain, around the Willaumez Peninsula, is linguistically a probable POC homeland (Ross 1989a). It is generally accepted by scholars that there is an association between the spread of Oceanic languages and the dispersal of Lapita pottery and associated artefacts from western Melanesia into the central Pacific (Shutler – Marck 1975, Spriggs 1984, Pawley – Green 1984), and this gives a clue as to why Pre-POC speakers might have settled at a New Britain location. One of the more important materials in Lapita culture was obsidian (volcanic glass), used to make cutting tools, and the Willaumez Peninsula with its string of volcanoes provides one of the most important sources of obsidian in the Indo-Pacific region: it was traded widely by Lapita potters — our putative Oceanic speakers — into the Pacific, and has also been found as far west as Borneo (Bellwood 1989). The association between POC and the Lapita bearers also helps us to date POC: Lapita pottery appeared in the Bismarck Archipelago sometime between 1900 and 1500 B.C. (Spriggs 1990).

POC speakers must have remained together long enough for the innovations listed above to have occurred in their speech. However, this need not have been more than a century or two. Linguistic and archaeological evidence indicates that from about 1250 B.C. Oceanic speakers spread rapidly across most of island Melanesia and into western Polynesia and Micronesia (Spriggs 1990). POC is shown in Figure 10 as dividing into three branches: Proto-Admiralty, Western Oceanic languages, and Proto-Central/Eastern-Oceanic. However, this division is a controversial simplification, to which we return below.

Again it is useful to put together a listing of groups as a basis for discussion:

- 1) Admiralty Islands (14)
- 2) St Matthias Islands (14)
- 3) Western Oceanic (Papua New Guinea and the western Solomon Islands) (6-10, 13-15)
- 4) Sarmi/Jayapura Bay (Irian Jaya) (3: *Yotafa-Sarmi group*)
- 5) Southeast Solomonic (15: *Gela-Guadalcanal and Malaita-San Cristobal groups*)
- 6) Utupua and Vanikoro (Te Motu Province, Solomon Islands) (15)
- 7) North/Central Vanuatu (16: *North and Central New Hebrides, East Santo, and Malekula Interior groups*)
- 8) South Vanuatu (16: *Erromanga, Tanna and Aneityum groups*)
- 9) Southern Oceanic (New Caledonia and the Loyalty Islands) (17)
- 10) Nuclear Micronesian (18)
- 11) Central Pacific (Rotuma, Fiji, Polynesia, New Zealand) (19)
- 12) Yapese (18)

Groups 1 and 2, the *Admiralties family* and the *St Matthias group*, are each clearly defined by a set of innovations, and Proto-Admiralty has been reconstructed (Blust 1978, Ross 1988: chapter 9). The St Matthias group may yet prove to be specifically associated with the Admiralties family; it is not closely relative to its southern neighbours on New Ireland. The Admiralties family is represented here by a word list from Nyindrou.

The *Western Oceanic* grouping, Group 3, was proposed by Ross (1988: 386-389). Because so little lexical material has been published for these 220 or so languages, they are represented here by fifteen word lists. With a few exceptions, POC **r* and **R* have merged in these languages, and they share reflexes of the innovative disjunctive pronoun **idri[a]* ‘they’ (instead of POC *(*k*)*ira*). Western Oceanic languages also sporadically reflect other features which imply that the languages of the region are descended from a dialect linkage with a history separate from that of other Oceanic groups. Its internal structure is discussed below.

Two groups of Oceanic languages have been recognised in Irian Jaya (Group 4), the Jayapura Bay and Sarmi Coast groups. The latter was the subject of a preliminary investigation by Grace (1971) and I have analysed the skimpy material for the other, the Jayapura Bay group.³⁵ It is sufficient to indicate that the two form a single higher-order *Sarmi/Jayapura Bay group*. Grounds for this grouping are (a) that POC **r*, **R* and **I* are merged in all member languages, and (b) that all member languages have a voiceless reflex of POC **b*.

Group 5, the *Southeast Solomonic* family is well defined by innovations and has been the subject of several pieces of reconstruction (Pawley 1972, Levy 1979, 1980, Tryon – Hackman 1983, Lichtenberk 1988). Its apparent conservatism makes it valuable for Oceanic comparative studies, and two word lists from the family are given in this work: Lau and Kwaio.

The languages of the tiny islands of *Utupua* and *Vanikoro* (Group 6) have been documented by Tryon – Hackman (1983), but have yet to be linked directly to any other group. Their speakers share their island group with speakers of the easternmost Papuan languages.

The languages of Group 7, *North/Central Vanuatu*, show innovations none of which is shared by all members of the group. Documented by Pawley (1972), Tryon (1976) and Clark (1985), the last-named describes the group as the outcome of a dialect linkage which has gradually differentiated out into local languages. He places an important boundary between North and Central Vanuatu, "with the boundary running between Santo and Malekula and between Raga and the remainder of Pentecost" (1985: 221). North/Central Vanuatu languages for which word lists appear in this work are Raga, Paama, Lewo and Port Sandwich.

The languages of the *South Vanuatu* group (Group 8) are spoken on the islands of Erromanga, Tanna and Aneityum, and extensive reconstruction of Proto-South Vanuatu and of interstage proto-languages has been done by Lynch (1978, 1983, 1986, 1990). They are represented here by North Tanna and Kwameria word lists. Clark (1985: 219-220) has also presented innovations shared by some languages in Groups 7 and 8, suggesting their descent from an ancestral linkage.

Group 9, the *Southern Oceanic* group includes all the languages of New Caledonia and the Loyalty Islands: its shared innovations were noted by Geraghty (1989). Representative word lists in this volume are provided for Nemi, Cèmuhî, A'jié, and Xârâcùù, on New Caledonia, and for the Loyalty Islands language Nengone. Geraghty also suggests that Groups 8 and 9 form a larger group, and implies that 9-10 may form a yet larger linkage. These hypotheses are circumstantially reasonable, but need more research.

The *Nuclear Micronesian* group (Group 10) covers all of Micronesia except Belau and part of the Marianas (see Western Malayo-Polynesian Group 5 above), and probably Yap. Group 10 is documented by Bender (1971) and Bender – Wang (1985), and Proto-Micronesian reconstruction has been done by Jackson (1983), who also provides a set of shared innovations. Although long considered an isolate, it is now considered that Nauruan probably also belongs to this group (Nathan 1973). It is represented

in the present work by word lists from Kiribati, Marshallese, Ponapean and Woleaian.

The integrity of the *Central Pacific* linkage (Group 11), covering a vast area of the Pacific, has long been recognised (Grace 1959, Pawley 1972), and its shared innovations and history (to be discussed below) have been reconstructed by Geraghty (1983, 1986, 1989).³⁶ Its representatives among the word lists here are Rotuman, West Fijian, East Fijian and, from the Polynesian subgroup of Central Pacific, Tongan, Samoan, Mele-Fila, Tahitian and Rapanui.

Group 12 consists of only one language, *Yapese*, which is also located on a Micronesian Island. Scholarly opinion is that it is not a member of the Nuclear Micronesian group, but is an Oceanic language whose exact position in the Oceanic group requires investigation.

The Oceanic group is perhaps the best understood of all the major groups in the Austronesian family in terms of the comparative method, and its areas of controversy are perhaps rather insignificant in comparison with those of other parts of the family.³⁷ Probably most Oceanic linguists now agree with Pawley's (1981) view that Oceania was settled quite rapidly from a POC homeland in the Bismarck Archipelago. The first Lapita assemblages in several regions of Remote Oceania (i.e. Oceania east of the Solomons) are all at around 1100-1000 B.C. (Spriggs 1990). The area of potential controversy among linguists concerns the details of the settlement process, and whether, because of the speed of settlement, we should regard Melanesia (the area bounded by Papua New Guinea, New Caledonia, and Fiji) as one great linkage (as Pawley 1981 suggests), within which only fairly localised groups of languages, each descended from a local interstage language, are recognisable, or whether, as we have done in Figure 10, we can treat the languages of the region as a set of subgroups from Proto-Oceanic to Proto-Central Pacific.

The truth doubtless lies somewhere between these two extremes. Linguistically, the integrity of the Admiralty Islands group is not in doubt, and the islands were almost certainly settled from the POC homeland at an early date (assuming that it was not the homeland: this possibility is remote, but not to be excluded). At about the same time, there was an eastward migration or migrations into the region represented in Figure 10 by the "Proto-Central/Eastern Oceanic" node, leaving behind in the Bismarck Archipelago speakers of dialects ancestral to the Western Oceanic languages. Although the dispersal of the latter was probably rather slower than the dispersal of Austronesian speakers into Remote Oceania, it is convenient to discuss the Western Oceanic languages first.

Although there are innovations which have affected most Western Oceanic languages, the fact that not all are affected suggests that this is another "stay-at-home" group, descended perhaps from a geographically fairly small linkage, which spread out rather more slowly than its departed sisters. The reasons for this tardiness are almost certainly associated with the fact that the mainland and at least the larger islands of Papua New Guinea had long been occupied by Papuan speakers, and in most areas of the mainland, Austronesian speakers have gained little more than a toehold on the coast.

Western Oceanic speakers spread first in two directions, to the east and to the west of the Willaumez Peninsula. To the east, they settled the north coast of New Britain coast and the islands of New Ireland (apparently from south to north), Bougainville, Choiseul, New Georgia, Santa Ysabel and their outliers to form the *Meso-Melanesian cluster*. Details of this dispersion are given by Ross (1988). The Meso-Melanesian languages share a set of innovations, setting them apart from the offshoots of Western Oceanic described in the next paragraph. It is probable either that Meso-Melanesian speakers encroached on territory already occupied by "Central/Eastern Oceanic" speakers or vice versa, as there is a sharp line dividing the Meso-Melanesian and South-East Solomonic groups from each other in the middle of the Solomons. The Meso-Melanesian cluster is represented here by word lists from Tolai, Roviana and Maringe.

To the west, Western Oceanic speakers settled the New Britain coast, and occupied islands in the Vitiaz Strait and close to the New Guinea mainland around the Huon Peninsula. Some speakers found their way into south-east Papua, forming the *Papuan Tip cluster*, whilst others later migrated along the north coast of Papua New Guinea, their dialects forming the *North New Guinea cluster*. The separate shared history of the two clusters is reflected in some exclusively shared lexical innovations. The migration to south-east Papua seems to have entailed initially a single settlement rather than a diffusion into the region, as the Papuan-Tip cluster is defined by a set of shared innovations and Proto-Papuan-Tip is readily reconstructible (Ross 1988, chapter 6). After settlement in the south-east Papuan region, speakers of a Papuan Tip language moved westwards along the south coast of Papua. This last movement can be correlated with the appearance of a pottery-bearing culture in the Central Province around 2000 years ago (Vanderwal 1973, Allen 1977a, 1977b, Bulmer 1982).

The North New Guinea cluster is characterised by the overlapping pattern of innovations which defines a linkage, but by no innovation defining the whole group (Ross 1988: 120). It is probably the outcome of a "stay-at-home" linkage which has diffused in fits and starts westward along the north

coast of Papua New Guinea, into the Huon Gulf, and around the south coast of New Britain. Circumstantially, at least, it is likely that the languages of the Sarmi/Jayapura Bay group are an extension of the north coast limb of the New Guinea cluster, and the limited evidence available tends to confirm this.³⁸ The earliest archaeological dates which can be confidently associated with North New Guinea speakers (on the New Britain coast west of the Willaumez Peninsula, and on the islands and the mainland coast of the Vitiaz Strait) are only about 1500 years ago (Lilley 1988, 1990).

Eight word lists are provided here from the North New Guinea cluster — Manam, Takia, Dami, Mangap-Mbula, Yabem, Kaulong, Buang and Adzera — and four from the Papuan Tip cluster — Kilivila, Tawala, Motu and Mekeo.

The eastward migration(s) into the Pacific shown in Figure 10 as the third branch of Proto-Oceanic are there labelled "Proto-Central/Eastern-Oceanic". However, neither this nor the next right-hand branch, "Proto-Remote Oceanic", has been satisfactorily defined in terms of shared innovations (although the next, Proto-Central-Pacific, has the characteristics of a linkage). The Central/Eastern Oceanic grouping, consisting of Groups 5-11, was proposed by Tryon — Lynch (1983) and is based on shared morphological innovations. However, reconstruction of POC morphology is as yet insufficient to be certain that these are innovations, rather than shared inheritances (Ross 1988: 393). Various other attempts have been made to define a large Oceanic subgroup in Eastern Oceania. In the published version of their paper, Tryon — Lynch (1985) exclude Southern Oceanic because the evidence for its inclusion seemed weak. Pawley's (1972) "Eastern Oceanic" included Groups 5, 7, 10 and 11, but for similar reasons was later withdrawn in favour of a "Remote Oceanic" which excluded South-East Solomonic (1977). This leaves us uncertain about the right branch(es) of Figure 10 between POC and Proto-Central-Pacific.

The question which underlies these subgrouping attempts is whether eastern Oceania was settled by a straightforward series of eastward movements of Oceanic speakers (which somewhere in Vanuatu would presumably have forked into Southern Oceanic and Central Pacific) or whether there was a more complex pattern of settlement involving multiple migrations (and perhaps back migrations) in various areas.

The definition of such a large subgroup in eastern Oceania cannot avoid evaluation of one innovation: loss of POC *R. Geraghty (1990) re-examines this loss and suggests a widespread pattern, namely that across much of the region which he labels "Eastern Oceanic" (coterminous with groups 5 to 11), POC *R is lost in a growing number of words, the further one moves

away from western Melanesia. If it could be shown that there is an implicational relationship whereby any word in which *R is lost at a given distance from western Melanesia is reflected without it in all more distant languages, then there would be grounds to posit a single, gradual movement of Oceanic speakers into the Pacific. However, this is not the hypothesis that Geraghty investigates, and since POC *R was evidently an unstable phoneme (probably a uvular trill), the possibility that its loss represents a series of independent parallel innovations is strong. It seems likely that growth in our understanding of larger-scale settlement patterns in eastern Oceania will depend on a number of pieces of detailed, localised research.

The presentation of the Central Pacific group in Figure 8 includes an interesting feature, namely that West Fijian and East Fijian, which many would assume to be variants of a single language, are not shown as each other's closest relatives. Rather, there is some evidence indicating that West Fijian is closer to Rotuman and East Fijian to the Polynesian languages. How has this arisen? Proto-Central-Pacific evidently was spoken in the Fijian islands from their settlement in 1200 B.C., where it diversified into a dialect linkage. This linkage became separated into two sections, a western and an eastern. Geraghty (1983) has labelled the eastern linkage "Proto-Tokalau-Fijian". Speakers of a western dialect reached Rotuma, and their speech became the ancestor of Rotuman (Pawley 1979). Speakers of a Tokalau-Fijian dialect found their way to one of the island groups east or north-east of Fiji sometime around 1000 B.C., where their language evolved into Proto-Polynesian (Geraghty 1983).³⁹ Since these departures, the stay-at-home western and Tokalau-Fijian linkages have through contact been resynthesised into the Fijian dialect network, within which western and eastern sections remain identifiable. Proto-Polynesian diversified into the languages of the Polynesian family. A final twist to this tale is that Polynesian speakers at least twice also settled on Rotuma, so that modern Rotuman has a very large Polynesian contribution to its lexicon (Biggs 1965).

Although they cover an enormous area, the Polynesian languages are linguistically one of the most coherent groups in Oceania, characterised by a clear set of phonological, morphosyntactic and lexical innovations. The languages are well documented and there is a fairly clearly defined internal subgrouping (Clark 1979). Extensive reconstruction has been carried out of Proto-Polynesian phonology (Dempwolff 1934, Elbert 1953, Biggs 1978), morphosyntax (Pawley 1966, 1967, 1970, Clark 1976, Wilson 1982), syntax (Hohepa 1969, Clark 1973, 1976, Chung 1978) and lexicon (Biggs – Walsh – Waqa 1970, Biggs 1979, n.d.).

External relationships of the Austronesian family

Over the years a number of proposals about the external relationships of the Austronesian languages — that is, about the closest relatives of PAN — have been made. The four most plausible proposals have to do, not unexpectedly, with the east Asian region. They are:

- a) the Japanese-Austronesian hypothesis, connecting Japanese with Austronesian;
- b) the Austro-Tai hypothesis, connecting Tai-Kadai and other languages of south China with Austronesian;
- c) the Sino-Austronesian hypothesis, connecting Old Chinese with Austronesian;
- d) the Austric hypothesis, connecting the Austro-Asiatic family with Austronesian.

In the case of Japanese, it is now generally accepted that putative Austronesian elements (if indeed they are Austronesian) are the results of borrowing or of a substratum left over after language shift rather than of genetic relationship, since neither sound correspondences nor shared grammatical morphemes have been found (Shibatani 1990: 103–109).

Before we turn to the other three proposals, it will be instructive to examine some internal Austronesian evidence about the origins of PAN. As we have seen, a large majority of reconstructed PAN items are disyllabic, with the form CVCV(C). Brandstetter (1906, 1916) drew attention to the fact that items with related meanings often share the same final syllable or "root", a matter which Blust (1988a) followed up in considerable detail (the term "root" is used in this section for the monosyllabic root, rather than as in the previous section for the PAN mono-, di- or trisyllable). For example, the root *pit 'press, squeeze together; narrow' is shared by PAN *ke(m)pit 'hold together, clamp', *Sa(m)pit 'hold together', PMP *ga(m)pit 'hold together', *qi(m)pit 'jammed, wedged close together', *le(m)pit 'fold', *li(m)pit 'press between two flat surfaces', *sipit 'squeeze, pinch, narrow', and a number of other forms, including the reduplicated *pitpit 'pinch' (Blust 1988a: 140–141). As Blust shows, the initial syllables of such items (in our example, *ke-, *Sa-, *ga-, *qi-, *le-, *li-, and *si-) also constitute an open class (1988a: 34–35). Since both elements in the many segmentable disyllabic reconstructions thus represent open classes, it is reasonable to suggest that these items may result from the compounding of monosyllabic roots, probably at some pre-PAN stage. But Blust points out that if this were so, we would also expect to be able to reconstruct the meanings of the initial

elements, and this he has been unable to do (1988a: 53). Work building on Blust's, however, suggests that some initial elements may be prefixes (Zorc 1990: 182) whilst others do have reconstructible meanings (Nothofer 1990). If this proves to be more widely the case, then the compounding hypothesis will be better supported than it is at present.

If PAN should prove to contain a large number of disyllabic items derived by the compounding of monosyllables, then an inference can be made about the typology of Pre-PAN, namely that it was monosyllabic and tonal. When a language with mainly monosyllabic roots forms a large number of disyllabic compounds which are then fossilised into new single-morpheme items, this usually happens because the language at its monosyllabic stage was tonal, and loss of tone is resulting in homophones. Compounds are then formed to compensate for homophony, and these compounds finally become unsegmentable items, as perhaps in PAN. The inference that pre-PAN may have been a mainly monosyllabic tone language is not particularly surprising, since such languages (belonging to several language families) occupy much of the region comprising China and mainland south-east Asia. It must be admitted, however, that we do not know when or where the monosyllabic tonal type originated in this region.

The proposal that PAN and the Tai-Kadai family are genetically related was initially made by Benedict (1942). The Tai-Kadai languages are scattered across southern China and cut a swathe down the centre of Indo-China, their best known member being Thai ("Tai" is used for the group of languages to which Thai belongs). In later papers, re-published together as Benedict (1975), he included the Ong-Be language and the Kam-Sui and Miao-Yao language groups in this macro-group, and named it "Austro-Thai", today usually spelled "Austro-Tai".⁴⁰ The languages of these groups and the Tai-Kadai languages are generally monosyllabic tone languages, and Benedict set out to show that monosyllabic roots in Tai-Kadai languages show sound correspondences with the stressed syllable of reconstructed PAN items. Benedict's hypothesis has received little support from Austronesianists, largely because his methods of establishing sound correspondences have been considered too loose. However, Reid (1985) has extended and re-evaluated Benedict's data base and suggests "that the similarities we find are of such kinds and in such quantities that they are highly unlikely to be accidental, and probably point to a genetic relationship". An important direction for future research is the systematic reconstruction of the proto-language or languages of the non-Austronesian members of the putative Austro-Tai grouping, so that comparisons can be

made with a thoroughness similar to that involved in the reconstruction of PAN.

A proposal that Old Chinese is a close relative of PAN has been put forward by Sagart (1990), who lists sound correspondences between PAN roots and Old Chinese monosyllables (a form of Chinese spoken around 1000 B.C.) and derives both from an earlier Proto-Sino-Austronesian (PSAN). The following are typical comparisons:

PAN	Old Chinese ⁴¹
*seli	'young shoot'
*aliq	'move, change place'
*tulis	'stripes, writing'
*suliq	'sucker'
*uliq	'be in a position to'
*daliq	'cause, motive'
*ləi	'bamboo sprout'
*ləi?	'come to'
*ləih	'pencil the eyebrows'
*l-y-əi?	'the sucker tree'
*l-y-əi?	'using, by means of'
*l-y-əi?	'because'

Final *-? and *-h of the Old Chinese reconstructions are posited on the basis of Haudricourt's (1954) proposal that Middle Chinese rising and falling tones were the outcomes of Old Chinese *-? and *-h respectively (whilst level tone reflected their absence). Sagart suggests that Old Chinese *-? and *-h in turn correspond with PAN *-q and *-s and were derived from PSAN *-q and *-s.

The lower three Old Chinese reconstructions above have recourse to infixation (-y-), which Sagart suggests was a characteristic of PSAN.

The main strengths of Sagart's paper lie in his presentation of 106 regular comparisons and in his ability to account for the development of Middle Chinese tones from PSAN in a way which agrees with accepted theories of tonogenesis. One weakness lies in the fact that the semantic distance between the items compared is sometimes greater than desirable — but semantic identity can hardly be expected in comparisons between proto-languages allegedly spoken around 1000 B.C. (Old Chinese) and 3000 B.C. (PAN). Other weaknesses lie in the difficulties inherent in reconstructing Old Chinese (see Norman 1988: 42-48) and in the shortness of Chinese forms, which increases the chance of accidental resemblance.

Whereas it was inferred above from Blust's theory of the PAN "root" that PAN disyllables developed from Pre-PAN monosyllables, Sagart makes the opposite assumption, proposing that Old Chinese monosyllables developed from PSAN disyllables by deleting the first syllable of the latter. His evidence for this lies in dialects of Middle Chinese which allowed initial consonant clusters. He cites the comparison PAN *buLút 'fibre' and Old Chinese *l-y-ut 'writing brush', together with its Middle Chinese (seventh

century A.D.) outcome *yuət*. He then points out that several Middle Chinese dialects have *p*-initial forms, e.g. Yan *pyuət*, as evidence that Old Chinese (or an earlier stage of Chinese) must have had a form such as **pl-y-ut*, i.e. a form which retained a reflex of the PSAN initial consonant reflected in PAN **buLút*. From forms like this, Sagart argues that whilst mainstream Old Chinese acquired monosyllabism by deleting PSAN initial syllables, some dialects deleted only the vowel of the initial syllable, at least in some items.

Sagart's PSAN hypothesis appears promising, but more decisive evidence for or against requires the amassing of a larger body of comparisons between Old Chinese and PAN and also further research into the reconstruction of a Proto-Chinese which would take account of a wider range of Chinese dialects. As Sagart points out, support for his hypothesis would not necessarily contradict the claim that Chinese belongs to the Sino-Tibetan language family, which is probably "unassailable" (Norman 1988: 13). It would mean, however, that the Austronesian family was a subgroup of an extended Sino-Tibetan.

The Austro-Asiatic hypothesis has its origins in Keane (1880), but it was Schmidt (1906) who proposed an "Austro-Asiatic" grouping comprising the Austro-Asiatic and Austronesian families and who coined these three terms. From the language map of mainland south-east Asia (see Wurm – Hattori 1983, map 25), it seems probable that the whole of this region — and perhaps a much larger area — was once occupied by Austro-Asiatic speakers. However, Tai-Kadai and Sino-Tibetan speakers have cut swathes into the region from the north and Austronesian speakers have taken over the coastal regions of the Malay Peninsula, leaving Austro-Asiatic speakers in occupation of scattered areas of mainland south-east Asia. Further afield, Austro-Asiatic languages are spoken in parts of India.

Perhaps the most conservative Austro-Asiatic language is spoken in the Nicobar Islands, some 200 kilometres north-west of Sumatra, and it is this which shows the clearest signs of a relationship with Austronesian. Reid (forthcoming) shows that Nicobarese has morphemes which are similar in both form and function to the PAN verbal morphemes **pa-* 'causative', **-um-* 'actor pivot/agent nominaliser', **ka-* 'inchoative', **ma-* 'undergoer pivot, involuntary', **-in-* 'perfective/patient nominaliser', and **-a* 'actor pivot, irrealis', the PAN ligature **na*, and perhaps also to PAN noun phrase markers ("determiners"). Reid argues that these are most unlikely to be Austronesian borrowings and indicates evidence that they are reconstructible for Proto-Austro-Asiatic. If this proves to be so, then it will be hard to deny a relationship between Proto-Austro-Asiatic and PAN. To support the Austro-Asiatic hypothesis, however, research is needed to seek out cognate vocabulary in

Austro-Asiatic and Austronesian languages and to establish sound correspondences between the two proto-languages.

Study of the external relationships of PAN is clearly still in its infancy, but we should note that the Austro-Tai, Sino-Austronesian and Austro-Asiatic hypotheses are not necessarily incompatible. Archaeological evidence suggests the possibility that the Tai-Kadai, Miao-Yao, Austro-Asiatic and Austronesian language families may all ultimately have arisen out of the cultures which resulted from the domestication of rice in the Yangtze Delta and spread across southern China as far as northern Thailand between about 6500 and 3000 B.C. (Bellwood forthcoming). The speakers of PAN or Pre-PAN were rice cultivators who probably first reached Taiwan from south-eastern China between 4000 and 3000 B.C. The place of the Sino-Austronesian hypothesis in this scenario is less clear, since the ancestors of the Chinese came from north of the Yangtze and apparently adopted rice-growing later than this period.

Closing remarks

Because this chapter is an attempt to expose the strengths and weaknesses of Austronesian linguistics — to indicate where I believe some of the "cutting edges" of the field may be — it is in a number of respects quite subjective and personal. This is inevitable: a scholar's view of the needs for future research in his field is a product of his personal synthesis of that field and of the speculations which grow from it. Some of what is written here is certainly speculative and hopefully, provocative.

There are several areas which could have been included here but were not. One of these is sociolinguistic studies of situations involving Austronesian speakers. I did not write about this, because there is nothing archetypally Austronesian about Austronesian sociolinguistics. But such studies are certainly a prerequisite to a more thorough understanding of the linguistic history of the Austronesians. In this connection it is also clear that there are far too few detailed studies of single languages and of languages in contact — in short, too few data. We hope that the present volume will give some impetus to rectifying this situation.

A large portion of this essay has been taken up with a survey of Austronesian subgrouping — of what we understand and what we don't understand. This is a choice which reflects my interests, but it is also the area in which scholars in other disciplines are most prone to ask questions, and sometimes to believe there must be answers where in fact there are still none. One point that emerges from this survey is that much research is needed not

only by linguists but also by archaeologists: the comparative linguist can reconstruct a sequence of events, and sometimes he can locate them, but he can never date them.

Notes

1. I would like to thank Alexander Adelaar, Tom Dutton, Mary Ritchie Key, Bernd Nothofer, Andrew Pawley and Darrell Tryon for their comments on earlier drafts of this chapter. Of course, I am responsible for the use I have made of these. I have also benefitted considerably from discussions with Peter Bellwood, Adrian Clynes, Charles Grimes, Laurence Reid, and David Zorc and from letters from Robert Blust.
2. These terms are taken from Ross (1988: 7-8). Pawley – Green (1984) use the terms *network-breaking* and *radiation* in senses similar to our *dialect differentiation* and *separation*.
3. I use the term "Taiwan", rather than "Formosa", for the island, since this is its modern name, but retain conventional usage among Austronesian linguists by speaking of the island's Austronesian languages as "Formosan", since the term "Taiwanese" is used for the island's most widely spoken Chinese dialect.
4. Dyen (1956) documents massive Malay borrowing in the Ngaju Dayak language of Borneo.
5. Cases of this are hard to document among Austronesian languages, but see Ross (1991a) for a probable case.
6. To determine cognacy with reasonable certainty, the investigator needs to know what phonological innovations each of the languages in question has undergone. But this entails the application of the time-consuming comparative method, which lexicostatistics was designed to avoid.
7. Examples of cases among the Austronesian languages of western Melanesia where language A has its highest percentage of shared cognates with language B, whilst language A shares a number of innovations with some third language which language B does not share, are given by Ross (1988, chapter 1).
8. Gaps occur in the table where no cognate occurs: for example, the Toba Batak word for 'ear' is *pilgol*, a word which has replaced the expected form **taliga*.
9. If the proto-language under reconstruction is not at the apex of the genetic tree, evidence can be drawn from languages which are *not* descended from it. For example, if a decision were needed about whether to reconstruct two proto-phonemes or one in Proto-Oceanic (see Figure 2), we might examine languages of the putative South Halmahera/West New Guinea and Central Malayo-Polynesian groups. If all agreed in having two phonemes where some Oceanic languages had one, we could tentatively conclude that Proto-Oceanic had inherited two proto-phonemes and that some Oceanic languages shared a merger (and perhaps formed a subgroup). If all agreed in having one phoneme, then the alternative conclusion would be supported. Until much more is known about the external relationships of Austronesian languages, however, PAN is at the apex of the genetic tree, and external evidence is not available.
10. Some scholars have argued for a conditioned split, but this enjoys only limited currency among Austronesian scholars (see Wolff 1991 and Ross 1991b).
11. As noted, the Formosan languages Amis and Bunun also share in the merger of PAN *C and *t, and may therefore form a subgroup with PMP. This is a hypothesis which needs further research.
12. Exceptions are Harvey (1982), who writes *t and *ts, and Dahl (1973, 1981), who writes *t₁ and *t₂; both attribute the values [t] and [θ] to these proto-phonemes. Blust (1990a) writes *t and *C, but also attributes the values [t] and [θ] to them.
13. This problem was to some degree already present in Dempwolff's reconstructed system. His pairs *t'/*d' and *k'/g' both ostensibly represented palatal sounds (1934: 15-16), but the phonetic distinction between them was not made clear.
14. Dyen and Tsuchida make considerable use of subnumeration, but this is shown here only where it affects equivalence between orthographies. The convention 'd-d-d' refers to word-initial, -medial, and -final instances of the phoneme.
15. The most radical attempt to make sense of PAN phonology is Wolff's (1988) reconstruction. However, his assumptions about Austronesian subgrouping differ from those described here, with the result that he gives less weight to Formosan data than other scholars have done and arrives at conclusions both about the phonological system and about phonetic values which are quite different from Blust's (1990b) account and from the account given here.
16. Sources of the materials in Table 4 are Li (1981) for Proto-Atayal, Tsuchida (1964) and Li (1978) for Saisiyat, Li (1976) for Thao, Tsuchida (1976) for Saaroa, Li (1977) for Proto-Rukai, Ting (1978) and Tsuchida (1983) for Proto-Puyuma, Ho (1978) for Proto-Paiwan. Materials for other languages are available (e.g. in Tsuchida 1976, Mei 1982, Li 1987) but they add no additional information to that present in Table 4.
17. Data supporting this reconstruction are found in the word lists under 06.910 and 09.265. See also the discussion in Harvey (1982).
18. Although Tsuchida (1976) retains the PAN palatals, it is noteworthy that he has few examples which contain any palatal other than *Z.
19. Dahl (1981: 152) suggests tentatively that the value of *S₂ was [z], whilst Li (1985) suggests [ʃ], but neither suggestion is justified by the Formosan reflexes.
20. There is substantial evidence that sound changes do occur which are stopped before they have affected all the relevant words in the language. For general discussion, see Chen – Wang (1975). For an Austronesian example, see Lincoln (1973).
21. Maddieson's data show a three-way phonemic contrast between /q/, /ʔ/ and /h/ or /f/ in Kurdish (Indo-European), Arabic and Neo-Aramaic (Semitic), Sui (Kam-Tai), and in

- languages from a variety of Amerindian groups, as well as in the Formosan language Atayal (where present-day /ʔ/ does not, however, necessarily reflect PAN *ʔ?).
22. Dempwolff (1934) recognised only the form CV(C)CVC, positing no initial or final vowels and no vowel sequences. The modern reformulation is due to Dyen (1953).
 23. Egerod (1965, 1966) gives an account of the Atayal verb system, but in different terminology from mine.
 24. Starosta – Pawley – Reid (1982) present a revision and extension of proposals made by Pawley – Reid (1980).
 25. The hypothesis that all Austronesian languages outside Taiwan belong to a single subgroup was first proposed by Dahl (1973).
 26. Since the name of the Amis, who live on the south-west coast of Taiwan, seems to be derived from PAN *qámiS ‘north’, (see item 12.470 in the word lists, and Reid 1982: 212, who makes a similar point.) it is possible that they were given this name by the Malayo-Polynesian speakers of the islands to the south, who remembered the Amis as their stay-at-home relatives. However, there is currently no linguistic support for this.
 27. The reader is referred to Dahl (1976: 74-75, 101, 131; 1981: 101-105) and to Blust (1976) for discussion of this seemingly incomplete merger.
 28. Reid (1990) discusses Negrito influence in certain Philippine languages, but does not refer to the South Mindanao group.
 29. Reid proposes a re-definition of terminology: the conventionally accepted Malayo-Polynesian group would become “Extra-Formosan” and the nasal infixation group would become “Malayo-Polynesian”.
 30. Nothofer (1988) continues a controversy ranging from Nothofer (1975) through Blust (1981) and Adelaar (1985, 1988). Here, however, I have attempted only to summarise the most recently expressed views of the protagonists.
 31. Note that this is a different use of the term ‘Hesperonesian’ than that coined by Dyen (1965b) and used by Tsuchida (1976) and Zorc in various publications.
 32. I am indebted to Charles Grimes for these examples.
 33. These innovations differ in their formulation somewhat from Blust's not only because my reconstruction of PMP differs from his PAN, but also because I have checked data from a larger number of SHWNG languages and obtained somewhat different results.
 34. Grace's orthography also included *ns, but this is interpreted by Ross (1989b) as representing reflexes of *s which arose independently by lenition processes in daughter-languages after the break-up of POC. Hence there is no POC symbol for it in the orthography used here.
 35. This material consists of word lists in Galis (1955), Cowan (1952) and Voorhoeve (1975), and short unpublished word lists by members of the Summer Institute of Linguistics.
 36. Geraghty's work builds on foundations laid by others, chiefly Grace (1959), Biggs (1965), Pawley (1970, 1972), Hockett (1976) and Wilson (1982).

37. A more extensive account of the dispersal of Oceanic languages is given by Pawley and Ross (forthcoming).
38. Data and argument supporting this are contained in an unpublished manuscript by the writer.
39. Although on subgrouping grounds Tonga is a good candidate for the location where a Central Pacific dialect evolved into Proto-Polynesian, the fact that Samoa was settled at about the same time as Tonga causes some puzzlement. Proto-Polynesian is so rich in innovations that its immediate break-up is not tenable. Either Proto-Polynesian developed as a dialect chain embracing Tonga, Niuatoputapu, Niuafo'ou, Samoa and perhaps Futuna and Uvea, or (more probably) it developed in one island group and replaced Central Pacific dialects spoken in the others. (I am grateful to Andrew Pawley for bringing this point to my attention.)
40. In more recent work, Benedict's “Tai-Kadai” and Kam-Sui groups and the “Ong-Be” language are combined into a Kam-Tai grouping with three subgroups: Tai, Li (“Kadai”) and Kam-Sui. “Ong-Be” (spoken on Hainan Island) is known as Lingao or Vo Limkou, and is attributed to the Tai subgroup (Wang 1987). The Miao-Yao language is not considered to belong to this grouping (Sun 1987).
41. Sagart's transcriptions are here re-transcribed in accordance with the conventions of the present work.

Bibliography

Adelaar, K. Alexander

- | | |
|-------------|--|
| 1981 | “Reconstruction of Proto-Batak phonology”, in: Blust (ed.), 1-20. |
| 1983 | “Malay consonant harmony: an internal reconstruction”, <i>NUSA</i> 16: 57–67. |
| 1985 | Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology (PhD diss., University of Leiden) [To be published in <i>Pacific Linguistics</i>]. |
| 1988 | “More on Proto-Malayic”, in: Ahmad – Zain (eds.), 59-77. |
| 1991a | “New ideas on the early history of Malagasy”, in: Steinhauer (ed.), 1–22. |
| 1991b | “The classification of the Tamanic languages” in: Tryon and Dutton (eds.). |
| forthcoming | “Borneo as a cross-roads for comparative Austronesian linguistics” in: Bellwood, Fox and Tryon (eds.). |

- Ahmad, M.T. – Z.M. Zain
 1988 *Rekonstruksi dan cabang-cabang Bahasa Melayu induk.* [Reconstruction and branching from Proto-Malay] (Kuala Lumpur: Dewan Bahasa dan Pustaka).
- Allen, Jim
 1977a "Management of resources in prehistoric coastal Papua", in: Winslow (ed.), 35–44.
 1977b "Sea traffic, trade and expanding horizons", in: Allen, Golson and Jones (eds.), 387–417.
- Allen, Jim – Jack Golson – Rhys Jones (eds.)
 1977 *Sunda and Sahul: prehistoric studies in southeast Asia, Melanesia And Australia* (London: Academic Press).
- Allison, E. Joe
 1979 "Proto-Danaw: a comparative study of Maranaw, Magindanaw, and Iranun", *PL*, A-55: 53–112.
- Baldi, Philip (ed.)
 1990 *Linguistic change and reconstruction methodology* (Berlin/New York: Mouton de Gruyter).
- Bellwood, Peter S.
 1978 *Man's conquest of the Pacific: the prehistory of Southeast Asia and Oceania* (Auckland: Collins).
 1985 *Prehistory of the Indo-Malaysian archipelago* (Sydney: Academic Press).
 1989 "Archaeological investigations at Bukit Tengkorak and Segarong, southeastern Sabah", *Bulletin of the Indo-Pacific Prehistory Association* 9: 122–162.
 forthcoming "Austronesian prehistory in southeast Asia: homeland, exodus and transformation", in: Bellwood – Fox – Tryon (eds.).
- Bellwood, Peter – James Fox – Darrell Tryon (eds.)
 forthcoming *The Austronesians in history: common origins and diverse transformations*.
- Bender, Byron W.
 1971 "Micronesian languages", in: Sebeok (ed.), 426–465.
- Bender, Byron W. – Judith W. Wang
 1985 "The status of Proto-Micronesian", in: Pawley – Carrington (eds.), 53–92.
- Benedict, Paul K.
 1942 "Thai, Kadai and Indonesian: a new alignment in southeastern Asia", *AmA* 44: 576–601.
- 1975 *Austro-Thai language and culture: with a glossary of roots* (New Haven: HRAF Press).
- Berg, René van den
 1991 "Muna dialects and the Munic languages: towards a reconstruction", in: Harlow – Clark (eds.).
- Biggs, Bruce G.
 1965 "Direct and indirect inheritance in Rotuman", *Lingua* 14: 383–415.
 1978 "The history of Polynesian phonology", in: Wurm – Carrington (eds.), 691–716.
 1979 "Proto-Polynesian word list II", *WPDA* 53.
 n.d. POLLEX (computer print-outs, University of Auckland).
- Biggs, Bruce G.– D.S. Walsh – Jocelyn Waqa
 1970 *Proto-Polynesian reconstructions with English to Proto-Polynesian finder list (WPDA)*.
- Blust, Robert A.
 1974a "Eastern Austronesian: a note", *WPLUH* 6(4): 101–107.
 1974b The Proto-North Sarawak vowel deletion hypothesis (unpublished Ph. D. dissertation, University of Hawaii).
 1976 Review of O.C. Dahl, *Proto-Austronesian*, *Lg* 52: 221–237.
 1977 "The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report", *WPLUH* 9/2: 1–15.
 1978 "Eastern Malayo-Polynesian: a subgrouping argument", in: Wurm – Carrington (eds.), 181–234 (Summary of the 130pp paper presented at the conference).
 1979 *The Proto-Oceanic palatals* (JPS Monograph 43).
 1981a "The reconstruction of Proto-Malayo-Javanic: an appreciation" [Review of Nothofer 1975], *BKI* 137: 456–469.
 1981b Variation in retention rate among Austronesian languages (Paper presented to the Third International Conference on Austronesian Linguistics, Bali, Indonesia).
 1982a "The linguistic value of the Wallace line", *BKI* 138/2-3: 231–250.
 1982b "An overlooked feature of Malay historical phonology", *BSOAS* 45: 284–299.
 1984 "More on the position of the languages of eastern Indonesia", *OL* 22–23: 1–28.
 1985 "The Austronesian homeland: a linguistic perspective", *Asian Perspectives* 26/1: 45–67.

- 1988a *Austronesian root theory: an essay on the limits of morphology* (Amsterdam/Philadelphia: Benjamins).
- 1988b "Malay historical linguistics: a progress report", in: Ahmad and Zain (eds.), 1-33.
- 1990a Central and Central-Eastern Malayo-Polynesian (Unpublished paper).
- 1990b "Patterns of sound change in the Austronesian languages", in: Baldi (ed.), 231-263.
- 1991 "On speech strata in Tiruray", *PL*, A-82.
- Blust, Robert A. (ed.)
- 1981 *Historical linguistics in Indonesia*, Part I. *NUSA* 10.
- 1991 *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace* (*PL*, C-117).
- Brandstetter, Renward
- 1906 *Ein Prodromus zu einem vergleichenden Wörterbuch der malaio-polynesischen Sprachen für Sprachforscher und Ethnographen* (Luzern: E. Haag).
- 1916 *An introduction to Indonesian linguistics*, translated by C.O. Blagden (London: The Royal Asiatic Society).
- Bulmer, Susan
- 1982 "West of Bootless Inlet: archaeological evidence for prehistoric trade in the Port Moresby area and the origins of the *hiri*", in: Dutton (ed.), 117-130.
- Capell, Arthur
- 1943 *The linguistic position of south-eastern Papua* (Sydney: Australasian Medical Publishing Co.).
- 1976 "Austronesian and Papuan 'mixed' languages: general remarks", in: Wurm (ed.), 527-579.
- Charles, Matthew
- 1974 "Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages", *OL* 13: 457-509.
- Chen, Matthew Y. - William S.-Y. Wang
- 1975 "Sound change: actuation and implementation", *Language* 51: 255-281.
- Chung, Sandra
- 1978 *Case marking and grammatical relations in Polynesian*. (Austin, Texas: University of Texas Press).

Clark, Ross

- 1973 "Transitivity and case in Eastern Oceanic languages.", *OL* 12: 559-605.
- 1976 *Aspects of Proto-Polynesian syntax* (1973 PhD diss., University of California, San Diego) (*Te Reo Monograph*) (Auckland: Linguistic Society of New Zealand).
- 1979 "Language", in: Jesse Jennings (ed.) *The prehistory of Polynesia* (Cambridge, Mass.: Harvard UP).
- 1985 "Languages of north and central Vanuatu: groups, chains, clusters and waves", in: Pawley - Carrington (eds.), 199-236.
- Clements, George N. - Samuel Jay Keyser
- CV phonology: a generative theory of the syllable* (Cambridge, Mass.: MIT Press).
- Clynes, Adrian J.
- 1990 Active articulator phonological patterning in Austronesian languages (Unpublished paper).
- Collins, James T.
- 1981 "Preliminary notes on Proto-West Central Maluku: Buru, Sula, Taliabo, and Ambelau", in: Blust (ed.), 31-45.
- 1982 "Linguistic research in Maluku: a report of recent field work", *OL* 21: 73-146.
- 1983 *The historical relationships of the languages of Central Maluku, Indonesia* (*PL*, D-47).
- Cooreman, Ann M. - Barbara Fox - Talmy Givón
- 1984 "The discourse definition of ergativity", *Studies in Language* 8: 1-34.
- Cowan, H.J.K.
- 1948 "Aanteekeningen betreffende de verhouding van het Atjéhsch tot de Mon-Khmer-talen" [Notes concerning the relationship of Acehnese and the Mon-Khmer languages], *BKI* 104: 429-514.
- 1953 "De austronesisch-papoea'se taalgrens in de onderafdeling Hollandia (Nieuw-Guinea)" [Austronesian-Papuan language boundaries in the Hollandia area], *Tijdschrift "Nieuw-Guinea"* 13: 133-143, 161-177, 201-206 (The Hague).
- 1974 "Evidence of long vowels in early Acehnese", *OL* 13: 187-212.
- 1991 "Achehnese dialects in connection with Chamic languages", in: Harlow - Clark (eds.).
- Crowley, Terry
- 1985 "Common noun phrase marking in Proto-Oceanic", *OL* 24: 135-193.

- Dahl, Otto Christian
- 1951 *Malgache et Maanjan: une comparaison linguistique* (Oslo: Egede-Instituttet).
 - 1976 *Proto-Austronesian* (London: Curzon Press) (2nd, revised, edition; 1st 1973, Oslo).
 - 1977 "La subdivision de la famille barito et la place du malgache", *Acta Orientalia* (Copenhagen) 38: 77–134.
 - 1981 *Early phonetic and phonemic changes in Austronesian* (Oslo: Universitetsforlaget).
- Davidson, Jeremy H.C.S. (ed.)
- 1990 *Pacific Island languages: essays in honour of G.B. Milner* (London: School of Oriental and African Studies/Honolulu: University of Hawaii Press).
- Dempwolff, Otto
- 1934 *Vergleichende Lautlehre des austronesischen Wortschatzes, 1: Induktiver Aufbau einer indonesischen Ursprache* (ZES Beiheft 15).
 - 1937 *Vergleichende Lautlehre des austronesischen Wortschatzes, 2: Deduktive Anwendung des Urindonesischen auf austronesischen Einzelsprachen* (ZES Beiheft 17).
 - 1938 *Vergleichende Lautlehre des austronesischen Wortschatzes, 3: Austronesisches Wörterverzeichnis* (ZES Beiheft 19).
- Durie, Mark
- 1989 "Proto-Chamic and Acehnese mid-vowels: towards Proto-Aceh-Chamic", *BSOAS* 52/4.
- Dutton, Thomas E.
- 1982a "Borrowing in Austronesian and non-Austronesian languages of coastal south-east mainland Papua New Guinea", in: Halim – Carrington – Wurm (eds.) (1982a), 109–177.
- Dutton, Thomas E. (ed.)
- 1982 *The hiri in history: further aspects of long-distance Motu trade in central Papua*. (*Pacific Research Monograph No.8.*) (Canberra: Australian National University).
- Dyen, Isidore
- 1953 *The Proto-Malayo-Polynesian laryngeals* (Baltimore: Linguistic Society of America).
 - 1956 "The Ngadju-Dayak 'old speech stratum'", *Lg* 32: 83–87.
 - 1965a "Formosan evidence for some new Proto-Austronesian phonemes", *Lingua* 14: 285–305.
- Egerod, Søren
- 1965 "Verb inflection in Atayal", *Lingua* 15: 251–281.
 - 1966 "Word order and word classes in Atayal", *Lg* 42: 346–369.
- Elbert, Samuel H.
- 1953 "Internal relationships of Polynesian languages and dialects", *SJA* 9: 147–173.
- Elkins, Richard E.
- 1974 "A Proto-Manobo word list", *OL* 13: 601–641.
 - 1984 "An extended Proto-Manobo word list", *PJL* 14/2-15/1: 218–229.
- Esser, S.J.
- 1938 "Talen [Languages]", in: *Atlas van tropisch Nederland*, 9–9b (Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap).
- Ferrell, Raleigh J.
- 1972 "Verb systems in Formosan languages", in: Jacqueline M.C. Thomas – Lucien Bernot (eds.) *Langues et techniques...I: Approche linguistique*, 121–128 (Paris: Klincksieck).
 - 1980 "Phonological subgrouping of Formosan languages", in: Naylor (ed.), 241–254.
 - 1982 *Paiwan dictionary* (PL, C-73).
- Fleischman, Eric
- 1981 "The Danao languages: Magindanao, Iranun, Maranao, and Ilanun", *PJL* 12: 57–78.
- Foley, William A. – Robert D. Van Valin
- 1984 *Functional syntax and universal grammar* (New York: CUP).
- Galis, K.W.
- 1955 "Talen en dialecten van Nederlands Nieuw Guinea", *Tijdschrift Nieuw Guinea* 16: 109–118, 134–135, 161–178.
- Gallman, Andrew F.
- 1979 "Proto-South-East Mindanao and its internal relationships", *PL*, A-55, 1–52.
- Geraghty, Paul A.
- 1983 *The history of the Fijian languages* (OL Special Publication 19).
 - 1986 "The sound system of Proto-Central-Pacific", in: Geraghty – Carrington – Wurm (eds.) (1986b), 289–312.

- 1989 "The reconstruction of Proto-Southern Oceanic", in: Harlow – Hooper (eds.), 141–156.
- 1990 "Proto-Eastern Oceanic *R and its reflexes", in: Davidson (ed.), 51–93.
- Geraghty, Paul – Lois Carrington – S.A. Wurm (eds.)
- 1986b *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics (PL, C-94)*.
- Gonzalez, Andrew, (ed.)
- 1973 *Parangal kay Cecilio Lopez* [Festschrift Lopez] (Quezon City: Linguistic Society of the Philippines).
- Grace, George W.
- 1959 *The position of the Polynesian languages within the Austronesian (Malayo-Polynesian) language family (IUPAL, memoir 16, supplement to IJAL 25)*.
- 1966 "Austronesian lexicostatistical classification: a review article [of Dyen 1965]", *OL* 5: 13–31 [Comment by Dyen, and Hymes, follows].
- 1969 "A Proto-Oceanic finder list", *WPLUH* [1/2]: 39–84.
- 1971 "Notes on the phonological history of the Austronesian languages of the Sarmi coast", *OL* 10: 11–37.
- Green, Roger C. – M. Kelly (eds.)
- 1972 *Studies in Oceanic culture history*, 3 (Pacific Anthropological Records 13) (Honolulu: Bernice P. Bishop Museum).
- Hagège, Claude
- 1986 *La langue Palau: une curiosité typologique*. (München: Wilhelm Fink).
- Halim, Amran – Lois Carrington – Stephen A. Wurm (eds.)
- 1982a *Papers from the Third International Conference on Austronesian Linguistics*, vol.1: *Currents in Oceanic (PL, C-74)*.
- 1982b *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers (PL, C-75)*.
- Harlow, Ray – Ross Clark (eds.)
- 1991 *VICAL 2: Western Austronesian languages: papers from the Fifth International Conference on Austronesian Linguistics* (Auckland: Linguistic Society of New Zealand).
- Harlow, Ray – Robin Hooper (eds.)
- 1989 *VICAL 1: Oceanic languages: papers from the Fifth International Conference on Austronesian Linguistics* (Auckland: Linguistic Society of New Zealand).
- Harmon, Carol W.
- 1979 "Proto-Manobo pronouns and case marking particles", *PL*, A-55: 113–134.
- Harvey, Mark
- 1982 "Subgroups in Austronesian", in: Halim – Carrington – Wurm (eds.) (1982b), 47–99.
- Haudricourt, André G.
- 1954 "Comment reconstruire le chinois archaïque", *Word* 10: 351–364.
- Ho, Dah-an
- 1978 "A preliminary comparative study of five Paiwan dialects" [in Chinese], *BIHP* 49: 565–681.
- Hockett, Charles F.
- 1976 "The reconstruction of Proto Central Pacific", *AnL* 18: 187–235.
- Hogan, David W.
- 1988 *Urak Lawoi': basic structures and a dictionary (PL, C-109)*.
- Hohepa, Patrick W.
- 1969 "The accusative-to-ergative drift in Polynesian languages", *JPS* 78: 295–329.
- Hollyman, Jim — Andrew Pawley (eds.)
- 1981 *Studies in Pacific languages & cultures, in honour of Bruce Biggs* (Auckland: Linguistic Society of New Zealand).
- Hopper, Paul J.
- 1979a "Some discourse sources of ergativity", *WPLUH* 11.2.
- 1979b "Aspect and foregrounding in discourse", in: Talmy Givón (ed.) *Syntax and semantics*, 12: *Discourse and syntax*, 213–241 (New York: Academic Press).
- 1983 "Ergative, passive, and active in Malay narrative", in: F. Klein-Andreu (ed.) *Discourse perspectives in syntax*, 67–88 (New York: Academic Press).
- 1988 "How ergative is Malay?", in: McGinn (ed.), 441–454.
- Hudson, Alfred B.
- 1967 *The Barito isolects of Borneo* (Ithaca: Cornell University Southeast Asia Program).
- Jackson, Frederick H.
- 1983 The internal and external relationships of the Trukic languages of Micronesia (PhD diss., University of Hawaii).
- 1986 "On determining the external relationships of the Micronesian languages", in: Geraghty – Carrington – Wurm (eds.) (1986b), 201–238.

- Josephs, Lewis S.
- 1975 *Palauan reference grammar* (Honolulu: University of Hawaii Press).
- Keane, A.H.
- 1880 "On the relations of the Indo-Chinese and Inter-Oceanic races and languages", *Journal of the Anthropological Institute* 9: 254-289.
- Kirch, Patrick V.
- 1988 "The Talepakemalai Lapita site and Oceanic prehistory", *National Geographic Research* 4: 328-342.
- Laycock, Donald C. – Werner Winter (eds.)
- 1987 *A world of language: papers presented to Professor S.A. Wurm on his 65th birthday* (PL, C-100).
- Levy, Richard S.
- 1979 "The phonological history of the Bugotu-Nggelic languages and its implications for Eastern Oceanic", *OL* 18: 1-31.
- 1980 "Languages of the southeast Solomon Islands and the reconstruction of Proto-Eastern-Oceanic", in: Naylor (ed.), 213-225.
- Li, Paul Jen-kuei
- 1976 "Thao phonology", *BIHP* 47/2: 219-244.
- 1977 "The internal relationships of Rukai", *BIHP* 48/1: 1-92.
- 1978 "A comparative vocabulary of Saisiat dialects", *BIHP* 49: 133-199.
- 1980 "The phonological rules of Atayal dialects", *BIHP* 51/2: 349-405.
- 1981 "Reconstruction of proto-Atayalic phonology", *BIHP* 52/2: 235-301.
- 1985 "The position of Atayal in the Austronesian family", in: Pawley – Carrington (eds.), 257-280.
- 1987 "The preglottalised stops in Bunun", in: Laycock – Winter (eds.), 381-387.
- Lichtenberk, Frantisek
- 1985a "Possessive constructions in Oceanic languages and in Proto-Oceanic", in: Pawley – Carrington (eds.), 93-140.
- 1985b "Syntactic-category change in Oceanic languages", *OL* 24: 1-84.
- 1988 "The Cristobal-Malaitan subgroup of Southeast Solomonic", *OL* 27: 24-62.
- Lilley, Ian
- 1988 "Prehistoric exchange across the Vitiaz Strait, Papua New Guinea", *Current Anthropology* 29: 513-516.
- 1990 Final report: prehistoric settlement and trade in northwest New Britain, Papua New Guinea (unpublished MS).
- Lincoln, Peter C.
- 1973 "Some possible implications of POC *t as /l/ in Gedaged", *OL* 12: 279-293.
- Lynch, John
- 1978 "Proto-South Hebridean and Proto-Oceanic", in: Wurm – Carrington (eds.), 717-779.
- 1981 "Melanesian diversity and Polynesian homogeneity: the other side of the coin", *OL* 20/2: 95-129.
- 1983 "Preliminary remarks on Proto-Eromangan", in: Lynch (ed.), 193-220.
- 1986 "The Proto-South Vanuatu pronominal system", in: Geraghty – Carrington – Wurm (eds.) (1986b), 259-284.
- Lynch, John D. (ed.)
- 1983 *Studies in the languages of Eromango* (PL, C-79).
- Lynch, John D. – Darrell T. Tryon
- 1985 "Central-Eastern Oceanic: a subgrouping hypothesis", in: Pawley – Carrington (eds.), 31-52.
- McCune, Keith
- 1979 "Passive function and the Indonesian passive", *OL* 18: 119-169.
- McGinn, Richard (ed.)
- 1988 *Studies in Austronesian linguistics* (Athens, Ohio: Center for Southeast Asia Studies, Ohio University).
- McManus, E.G.
- 1977 *Palauan-English dictionary* (Honolulu: University of Hawaii Press).
- Maddieson, Ian
- 1984 *Patterns of sounds* (Cambridge: Cambridge University Press).
- Mei, Kuang
- 1982 "Pronouns and verb inflection in Kanakanavu", *Tsing Hua Journal of Chinese Studies* (New Series) 14: 207-231.
- Mester, R. Armin
- 1988 *Studies in tier structure* (New York: Garland).

- Mills, Roger F.
- 1975 Proto-South Sulawesi and Proto-Austronesian phonology (PhD diss., University of Michigan) (Ann Arbor: University Microfilms International).
- Naylor, Paz Buenaventura (ed.)
- 1980 *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages* (Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies).
- Norman, Jerry
- 1988 *Chinese* (Cambridge: Cambridge University Press).
- Nothofer, Bernd
- 1975 *The reconstruction of Proto-Malayo-Javanic* (VKI 73) (The Hague: Martinus Nijhoff).
- 1984 "Further evidence for the reconstruction of *-əy and *-əw", *BKI* 140: 451–458.
- 1986 "The Barrier Island languages in the Austronesian language family", in: Geraghty – Carrington – Wurm (eds.) (1986b), 87–109.
- 1988 "A discussion of two Austronesian subgroups: Proto-Malay and Proto-Malayic", in: Ahmad – Zain (eds.), 34–58.
- 1990 "Review article on Blust (1988)", *Oceanic Linguistics* 29: 132–152.
- 1991 "The relationship between the languages of the Barrier Islands and the Sulawesi-Philippine languages", in: Tryon – Dutton (eds.).
- PalleSEN, A. Kemp
- 1985 *Culture contact and language convergence* (Linguistic Society of the Philippines Monograph 24) (1977 PhD diss., University of California, Berkeley) (Manila: SIL).
- Pawley, Andrew K.
- 1966 "Samoan phrase structure: the morphology-syntax of a Western Polynesian language", *AnL* 8/5: 1–63.
- 1967 "The relationships of Polynesian outlier languages", *JPS* 76: 259–296.
- 1970 "Grammatical reconstruction and change in Polynesia and Fiji", in: Wurm – Laycock (eds.), 301–367.
- 1972 "On the internal relationships of eastern Oceanic languages", in: Green – Kelly (eds.), 1–142.
- 1973 "Some problems in Proto-Oceanic grammar", *OL* 12: 103–188.
- 1977 On redefining 'Eastern Oceanic' (Mimeo).
- 1979 *New evidence on the position of Rotuman* (WPDA 56).
- 1981 "Melanesian diversity and Polynesian homogeneity: a unified explanation for language", in: Hollyman — Pawley (eds.), 269–309.
- Pawley, Andrew – Lois Carrington (eds.)
- 1985 *Austronesian linguistics at the 15th Pacific Science Congress* (PL, C-88).
- Pawley, Andrew – Roger C. Green
- 1973 "Dating the dispersal of the Oceanic languages", *OL* 12: 1–67.
- 1984 "The Proto-Oceanic language community", *JPH* 19: 123–146 (reprinted in: Robert Kirk – Emöke Szathmary (eds.), 1985, *Out of Asia: peopling the Americas and the Pacific*, 161–184, Canberra: Department of Pacific History, ANU).
- Pawley, Andrew – Lawrence A. Reid
- 1980 "The evolution of transitive constructions in Austronesian", in: Naylor (ed.), 103–130.
- Pawley, Andrew – Malcolm D. Ross
- forthcoming "The prehistory of the Oceanic languages: a current view", in: Bellwood – Fox – Tryon (eds.).
- Pecoraro, Ferdinando
- 1979 *Éléments de grammaire Taroko* (Paris: Association Archipel).
- Ray, Sidney H.
- 1911 "Comparative notes on Maisin and other languages of eastern Papua", *JRAI* 41: 397–405.
- 1926 *A comparative study of the Melanesian island languages* (CUP).
- Reid, Lawrence A.
- 1982 "The demise of Proto-Philippines", in: Halim – Carrington – Wurm (eds.) (1982b), 201–216.
- 1985 "Benedict's Austro-Tai hypothesis – an evaluation", *Asian Perspectives* 26: 19–34.
- 1990 The search for original Negrito: Negrito languages as creolized Austronesian (Unpublished paper).
- forthcoming "The Nicobarese evidence for Austric", in: Bellwood – Fox – Tryon (eds.).
- Ross, Malcolm (M.D.)
- 1987 "A contact-induced morphosyntactic change in the Bel languages of Papua New Guinea", in: Laycock – Winter (eds.), 583–601.
- 1988 *Proto Oceanic and the Austronesian languages of western Melanesia* (PL, C-98).

- 1989a "Early Oceanic linguistic prehistory: a reassessment", *JPH* 24: 135–149.
- 1989b "Proto-Oceanic consonant grade and Milke's *nj", in: Harlow – Hooper (eds.), 433–496.
- 1991a "Areal phonological features in north central New Ireland", in: Tryon – Dutton (eds.).
- 1991b The sound of Proto Austronesian: an outsider's view of the Formosan evidence. (Paper presented to the Sixth International Conference on Austronesian Linguistics, Honolulu).
- 1991c "How conservative are sedentary languages? Evidence from western Melanesia", in: Blust (ed.), 433–451.
- Sagart, Laurent
- 1990 Chinese and Austronesian are genetically related (Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, Arlington, Texas).
- Schmidt, Wilhelm
- 1906 "Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens un Austronesiens", *Archiv für Anthropologie* 33: 59–109.
- Sebeok, Thomas A. (ed.)
- 1971 *Current trends in linguistics*, vol. 8: *Linguistics in Oceania* (The Hague: Mouton).
- Shibatani, Masayoshi
- 1990 *The languages of Japan* (CUP).
- Shutler, R. – J.C. Marck
- 1975 "On the dispersal of the Austronesian horticulturalists", *Archaeology and Physical Anthropology in Oceania* 10/2: 81–113.
- Sneddon, J.N.
- 1978 *Proto-Minahasan: phonology, morphology and wordlist* (PL, B-54).
- 1984 *Proto-Sangiric and the Sangiric languages* (PL, B-91).
- 1989 "The North Sulawesi microgroups: in search of higher level connections", in: Sneddon (ed.), 83–106.
- Sneddon, J.N., ed.
- 1989 *Studies in Sulawesi linguistics No.1*. NUSA.
- Sneddon, J.N. – Hunggu Tadjuddin Usup
- 1986 "Shared reflexes in the Gorontalic language group: implications for subgrouping", *BKI* 142: 407–426.
- Springgs, Matthew
- 1984 "The Lapita Cultural Complex", *Journal of Pacific History* 19: 204–205.
- 1990 "Dating Lapita: another view", in: Matthew Spriggs (ed.), *Lapita design, form and composition: proceedings of the Lapita Design Workshop*, 6–27 (*Occasional Papers in Prehistory* 19) (Canberra: Department of Prehistory, Research School of Pacific Studies, Australian National University).
- Starosta, Stanley – Louise Pagotto
- 1991 "The grammatical genealogy of Chamorro", in: Harlow – Clark (eds.).
- Starosta, Stanley – Andrew K. Pawley – Lawrence A. Reid
- 1982 "The evolution of focus in Austronesian", in: Halim – Carrington – Wurm (eds.) (1982b), 145–170 (An abridged version of the paper presented to the conference).
- Steinhauer, Hein (ed.)
- 1991 *Papers in Austronesian linguistics No.1* (PL, A-81).
- Stern, Robert H.
- 1975 "Sobei verb inflection", *OL* 14: 128–145.
- Stern, Joyce K.
- 1987 "Sobei verb morphology reanalyzed to reflect POC studies", *OL* 26: 30–54.
- Stokhof, W.A.L. (ed.), with Lia Saleh-Bronckhorst
- 1980 *Holle lists: vocabularies in languages of Indonesia*, vol.2: *Sula and Bacan Islands, north Halmahera, south and east Halmahera*. (PL, D-28).
- Sun Hongkai
- 1987 "Minority languages in southern China", in: Wurm, T'sou et al. (eds.), map C-6.
- Ting Pang-Hsin
- 1978 *Reconstruction of Proto-Puyuma phonology* (BIHP 49/3).
- Topping, Donald M.
- 1973 *Chamorro reference grammar* (PALI Language Texts: Micronesia) (Honolulu: UP of Hawaii).
- 1975 *Chamorro-English dictionary* (Honolulu: UP of Hawaii).
- Tryon, Darrell T.
- 1976 *New Hebrides languages: an internal classification* (PL, C-50).
- Tryon, Darrell T. – Thomas E. Dutton (eds.)
- 1992 [Language contact and change in the Austronesian world] (Berlin: Mouton De Gruyter).

- Tryon, Darrell T. – Brian D. Hackman
 1983 *Solomon Islands languages: an internal classification* (PL, C-72).
- Tsuchida, Shigeru
 1964 "Preliminary reports on Saisiyat: phonology", *Gengo Kenkyu* 46: 42-42.
- 1976 *Reconstruction of proto-Tsouic phonology* (Study of Languages and Cultures of Asia and Africa Monograph Series, 5) (Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa).
- 1980 "Puyuma (Tamalakaw dialect) vocabulary, with grammatical notes and texts" [in Japanese], in: *Kuroshio no minzoku, bunka, gengo* [The tribes, cultures and languages of the Black current], 183–307 (Tokyo).
- 1983 "Puyuma-English index", *Working Papers in Linguistics* 83: 10–63 (University of Tokyo).
- Uhlenbeck, Eugenius M.
 1949 *De structuur van het Javaanse morpheem* (VBG 78) (Bandoeng: Nix).
- 1950 "The structure of the Javanese morpheme", *Lingua* 2: 239–270.
- Usup, Hunggu Tadjuddin
 1986 Rekonstruksi Protobahasa Gorontalo-Mogondow (unpublished dissertation, University of Indonesia).
- Vanderwal, R.
 1973 Prehistoric studies in central coastal Papua (unpublished Ph.D. dissertation, Australian National University).
- Verheijen, Jilis A.J.
 1986 *The Sama/Bajau language in the Lesser Sunda Islands* (PL, D-70).
- Voorhoeve, C.L.
 1975 *Languages of Irian Jaya: checklist, preliminary classification, maps, wordlists* (PL, B-32).
- Walker, Dale F.
 1976 *A grammar of the Lampung language: the Pesisir dialect of Way Lima* (Revised version of his PhD diss., Cornell University) (Jakarta: NUSA).
- Wang Jun
 1987 "Kam-Tai languages", in: Wurm, T'sou et al. (eds.), map C-7.
- Weinreich, Uriel
 1963 *Languages in contact* (The Hague: Mouton).
- White, J. Peter – J. Allen – J. Specht
 1988 "The Lapita Homeland Project", *Australian Natural History* 22: 410-416.
- Wilson, William H.
 1982 *Proto-Polynesian possessive marking* (PL, B-85).
- Winslow, John H. (ed.)
 1977 *The Melanesian environment* (Canberra: Australian National University Press).
- Wolff, John U.
 1972 *A dictionary of Cebuano Visayan* (PJL Special Monograph 4) (Ithaca, New York: Southeast Asia Program, Cornell University, and Linguistic Society of the Philippines).
- 1973 "Verbal inflection in Proto-Austronesian", in: Gonzales (ed.), 71–91.
- 1974 "Proto-Austronesian *r and *d", *OL* 13: 77–121.
- 1988 "The PAN consonant system", in: McGinn (ed.) 1988: 125–147.
- 1991 "The Proto-Austronesian phoneme *t and the grouping of the Austronesian languages", in: Blust (ed.), 535–549.
- Wolff, J.U. – Soepomo Poedjosoedarmo
 1982 *Communicative codes in central Java* (Ithaca: Department of Asian Studies, Cornell University).
- Wouk, Fay
 1986 "Transitivity in Batak and Tagalog", *Studies in Language* 10: 391-424.
- Wurm, Stephen A. (ed.)
 1976 *New Guinea area languages and language study, vol.2: Austronesian languages* (PL, C-39).
- Wurm, Stephen A. – Lois Carrington (eds.)
 1978 *Second International Conference on Austronesian Linguistics: proceedings* (2 fascicles) (PL, C-61).
- Wurm, Stephen A. – Shirô Hattori (eds.)
 1981 *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).
- 1983 *Language atlas of the Pacific area, part II: Japan area, Taiwan (Formosa), Philippines, mainland and insular south-east Asia* (Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy; PL, C-67).
- Wurm, Stephen A. – Donald C. Laycock (eds.)
 1970 *Pacific linguistic studies in honour of Arthur Capell* (PL, C-13).

- Wurm, Stephen A. – B. T'sou – D. Bradley – Li Rong – Xiong Zhengui – Zhang Zhenxing – Fu Maoji – Wang Jun – Dob (eds.)
 1987 *Language atlas of China* (Hong Kong: Longman; PL, C-102).
- Wurm, Stephen A. - B. Wilson
 1975 *English finderlist of reconstructions in Austronesian languages (post-Brandstetter)* (PL, C-33).
- Zorc, R. David
 1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction* (PL, C-44).
 1978 "Proto-Philippine word accent: innovation or Proto-Hesperonesian retention?", in: Wurm – Carrington (eds.), 67–119.
 1982 "Where, o where, have the laryngeals gone? Austronesian laryngeals re-examined", in: Halim – Carrington – Wurm (eds.) (1982b), 111–144.
 1983 "Proto-Austronesian accent revisited", *PJL* 14/1: 1–24.
 1986 "The genetic relationships of Philippine languages", in: Geraghty – Carrington – Wurm (eds.) (1986b), 147–173.
 1990 "The Austronesian monosyllabic root, radical or phonestheme", in: Baldi (ed.), 175–194.

4. Listing of Austronesian languages

Barbara F. Grimes

Joseph E. Grimes

Malcolm D. Ross

Charles E. Grimes

Darrell T. Tryon

The inventory of Austronesian languages that follows is adapted from the Cornell-SIL Language Archive, the data base from which the 1988 *Ethnologue: languages of the world* (eleventh edition) was drawn. The Archive itself is the work of the first two authors: the last three present significant updates within that framework, particularly in the language classification.

Such a catalogue of languages necessarily depends on the quality and availability of the primary sources. The list represents interaction with primary researchers over a period of years. One aim has been to clarify what the primary researchers mean by relative terms such as "language", "dialect", "closely related", and "clearly distinct". Why have they chosen to use one particular name for a language in place of others used in the literature previously? Do their figures for number of speakers reflect the number of active users? the total number of those identified with the ethnic group regardless of whether or not a significant portion of the ethnic group have shifted completely to another language? first and second language speakers together?

The issues are varied and complex, and each language listed is a fascinating reflection of the sum of the linguistic, social, historical, political and geographical factors that belong to the speech community in question. In one area the social attitudes and political dynamics may force two speech communities that are linguistically close (in phonology, morpho-syntax, and lexicon) to be considered separate languages. In another area different social dynamics may create pressure to list as a single language two speech communities that are linguistically much more diverse and where intelligibility is marginally or significantly blocked. Or there may be extensive chaining effects that have simply not been sorted out well enough yet to give a partitioning of an area that reflects the actual communication potential of the speech varieties that are in use. The issues are complex, and the quest for an accurate picture is not made easier when one linguist's "dialect" is another linguist's "language".