

PACIFIC LINGUISTICS

Series C - No. 75

DEPT. OF LINGUISTICS
SCHOOL OF PACIFIC STUDIES

PAPERS FROM THE
THIRD INTERNATIONAL CONFERENCE ON
AUSTRONESIAN LINGUISTICS

*

Volume 2

TRACKING THE TRAVELLERS

Amran Halim
Lois Carrington
S.A. Wurm
eds



Department of Linguistics
Research School of Pacific Studies
THE AUSTRALIAN NATIONAL UNIVERSITY

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

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

EDITOR: S.A. Wurm

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

EDITORIAL ADVISERS:

B.W. Bender University of Hawaii	John Lynch University of Papua New Guinea
David Bradley La Trobe University	K.A. McElhanon University of Texas
A. Capell University of Sydney	H.P. McKaughan University of Hawaii
S.H. Elbert University of Hawaii	P. Mühlhäusler Linacre College, Oxford
K.J. Franklin Summer Institute of Linguistics	G.N. O'Grady University of Victoria, B.C.
W.W. Glover Summer Institute of Linguistics	A.K. Pawley University of Auckland
G.W. Grace University of Hawaii	K.L. Pike University of Michigan; Summer Institute of Linguistics
M.A.K. Halliday University of Sydney	E.C. Polomé University of Texas
A. Healey Summer Institute of Linguistics	Gillian Sankoff University of Pennsylvania
L.A. Hercus Australian National University	W.A.L. Stokhof National Center for Language Development, Jakarta;
Nguyễn Đăng Liêm University of Hawaii	University of Leiden
	J.W.M. Verhaar Gonzaga University, Spokane

The series of volumes C74, C75, C76, and C77 is published by agreement with
the Committee for International Conferences on Austronesian Linguistics.

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

The Secretary
PACIFIC LINGUISTICS
Department of Linguistics
Research School of Pacific Studies
The Australian National University
P.O. Box 4, Canberra, A.C.T. 2600
Australia.

Copyright © The Authors

First Published 1982

Typeset by Christine Billerwell

Printed by A.N.U. Printing Service

Covers by Patria Printers

Bound by Adriatic Bookbinders Pty. Ltd.

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

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

National Library of Australia Card Number and ISBN 0 85883 275 5

TABLE OF CONTENTS

		page
INTRODUCTION		iv
THE CONTRIBUTORS		v
LOCATION MAP		vi
Proto-Austronesian *c, *z, *g and *t	John U. WOLFF	1-30
The present status of some Austronesian subgrouping hypotheses	Isidore DYEN	31-35
A syntactic model for the comparative study of Austronesian languages	Teodoro A. LLAMZON	37-46
Subgroups in Austronesian	Mark HARVEY	47-99
Towards a typological reconstruction of the verbal system in Proto-Austronesian	J.C. ANCEAUX	101-110
Where, O where, have the laryngeals gone? Austronesian laryngeals re-examined	R. David ZORC	111-144
The evolution of focus in Austronesian	Stanley STAROSTA, Andrew K. PAWLEY and Lawrence A. REID	145-170
Atayalic final voiced stops	Paul Jen-kuei LI	171-185
Prothesis in the languages of Central Maluku: an argument from Proto-Austronesian grammar	James T. COLLINS	187-200
The demise of Proto-Philippines	Lawrence A. REID	201-216
The Halmahera connection: a case for prehistoric traffic through Torres Straits	C.L. VOORHOEVE	217-239
Sound changes in the Gorontalo language	J. NOORDUYN	241-261
The inadequacy of the inertial development principle in accounting for sound changes in several Austronesian languages	FARID M. ONN and Mangantar SIMANJUNTAK	263-284
Phonological problems of loanwords in Bahasa Indonesia	Hans LAPOLIWA	285-297
The phonology of Javanese vowels	Colin YALLOP	299-319
Synchronical description at the phonetic and syllabic level of Modang (Kalimantan Timur) in contrast to Kenyah, Kayan, and Palawan (Philippines)	Nicole REVEL-MACDONALD	321-331

INTRODUCTION

The Third International Conference on Austronesian Linguistics was held in Bali, Indonesia, in January 1981. Amran Halim, as Conference Organiser, and Stephen Wurm, as OCICAL committee member (and now its Chairman), and also as General Editor of Pacific Linguistics publications, decided that in addition to the Proceedings volume which would be produced in Jakarta, a selection of papers should be published, largely for the international readership, by Pacific Linguistics. This is the second in a series of modest volumes presenting a selection of the papers from the conference: it includes papers concerning the Western Austronesian area, tracking the comparative relationships of the languages of this area, following the interaction of language upon language over periods of time both great and small, tracking the genesis of these shifting, changing languages, and tracking phonological changes in the area. Volume 1 in the series, *Currents in Oceanic*, has already appeared; volume 3, which includes papers on sociolinguistics, and is titled *Accent on variety*, and volume 4, containing papers on syntax, morphology, etc. and titled *Thematic variation*, will follow shortly.

The editors wish to thank especially, for their helpful comments, Bert Voorhoeve, Laurie Reid, Alan Baxter, David Zorc, and John Wolff. As ever, our grateful thanks are due to Christine Billerwell for her thoughtful typesetting, and to Lio Pancini for his mapmaking skill.

AMRAN HALIM
Pusat Pembinaan dan Pengembangan Bahasa
Departemen Pendidikan dan Kebudayaan, Jakarta

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

S.A. WURM
Department of Linguistics
Research School of Pacific Studies
Australian National University, Canberra

THE CONTRIBUTORS

J.C. ANCEAUX is Professor of Austronesian Linguistics in the University of Leiden

JAMES T. COLLINS is a Lecturer in Historical Linguistics and Malay Dialectology at Universiti Kebangsaan Malaysia

ISIDORE DYEN is Professor of Comparative Linguistics and Austronesian Languages at Yale University

FARID M. ONN is Director of the Institute of Malay Language, Literature and Culture, and Associate Professor in Malay Studies and Linguistics at the Universiti Kebangsaan Malaysia

MARK HARVEY is a postgraduate student in the Department of Linguistics, Faculty of Arts, Australian National University

HANS LAPOLIWA is a consultant with the National Center for Language Development, Ministry of Education and Culture, Jakarta

PAUL JEN-KUEI LI is Research Fellow at the Institute of History and Philology, Academia Sinica, and Professor in Linguistics at the Department of Archaeology and Anthropology, National Taiwan University

TEODORO A. LLAMZON is a Specialist in General Linguistics at the Southeast Asian Ministers of Education (SEAMEO) Regional Language Centre in Singapore

J. NOORDUYN is Senior Research Fellow in the Department of Languages and Cultures of South-East Asia and Oceania at the University of Leiden

ANDREW K. PAWLEY is Associate Professor in the Department of Anthropology, University of Auckland

LAWRENCE A. REID is a Researcher in the University of Hawaii at Manoa with a split appointment between the Social Science Research Institute and the Department of Linguistics

NICOLE REVEL-MACDONALD is an Associate of the Centre du Documentation et de Recherches sur l'Asie du Sud-Est et le Monde Insulindien, C.N.R.S., Paris

MANGANTAR SIMANJUNTAK is Senior Lecturer in Psycho-Linguistics in the Universiti Kebangsaan Malaysia

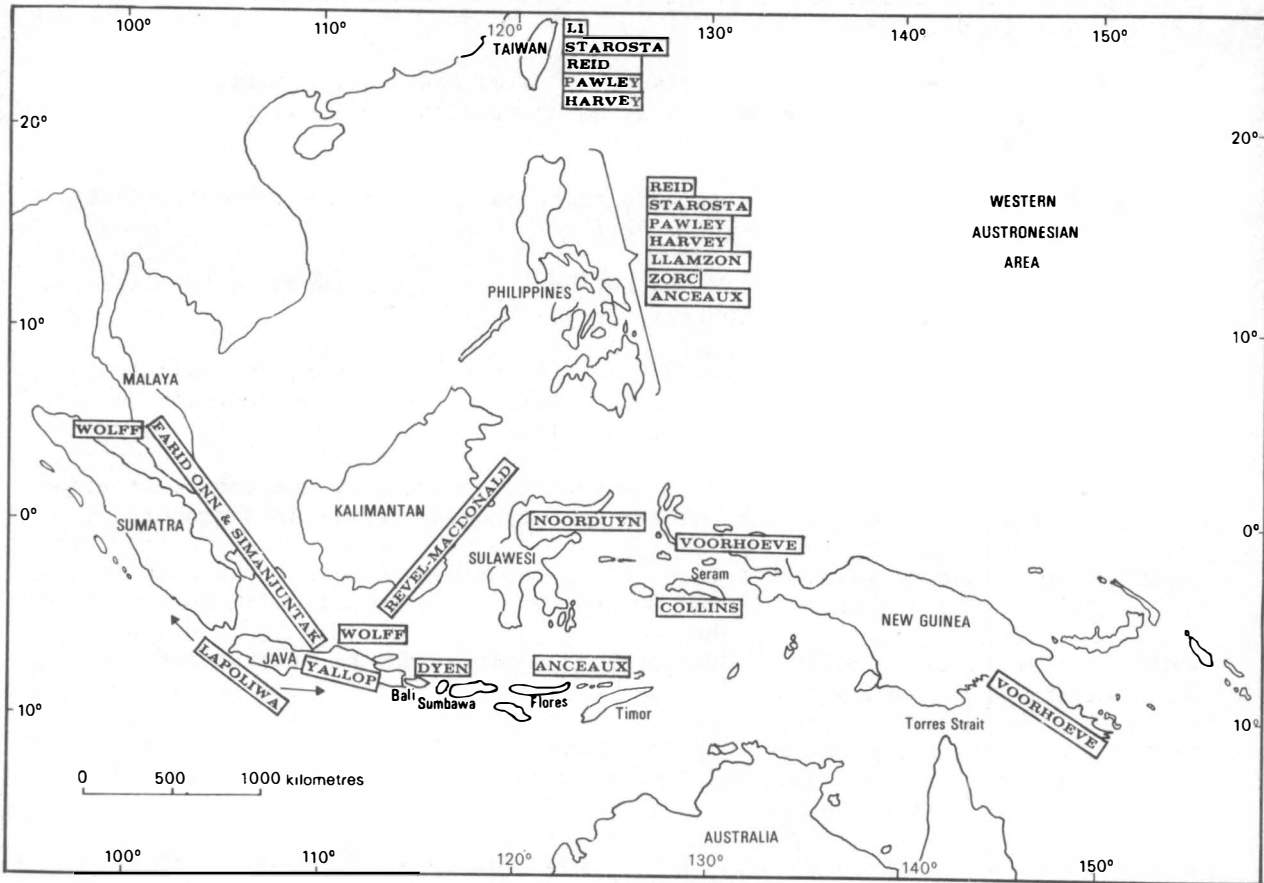
STANLEY STAROSTA is a Professor of Linguistics at the University of Hawaii at Manoa

C.L. VOORHOEVE is a Senior Fellow in the Department of Linguistics, Research School of Pacific Studies, Australian National University

JOHN U. WOLFF is Professor of Linguistics and Asian Studies at Cornell University

COLIN YALLOP is Senior Lecturer in Linguistics at Macquarie University in Sydney

R. DAVID ZORC is a Senior Lecturer in the School of Australian Linguistics of Darwin Community College, Northern Territory



Location of language areas referred to by the authors

PROTO-AUSTRONESIAN *c, *z, *g, AND *t

John U. Wolff

0. INTRODUCTION

Dempwolff in his Proto-Austronesian Phonology posited twenty-three consonantal phonemes including the semivowels *y* and *w*. Over the past thirty years a great deal of the work which has been done on Austronesian phonology has been to look for correspondences which are not taken care of by Dempwolff's phonemes. At the latest count I reckon an additional twenty to thirty consonant phonemes which have been posited by various people at various times, mostly in a very tentative way. This gives us a total of between forty-three and fifty-three consonants, an incredible number of consonants in a system which distinguishes only two modes of articulation, voiced and voiceless. This is further an incredible collection in view of the fact that the modern Austronesian languages, for the most part, have a very modest system of consonant phonemes. Something is clearly amiss, and I suggest that we reconsider the whole question of establishing the phonology of the proto-language. Although most of these phonemes were proposed on a tentative basis, scholars who have occasion to quote PAN forms containing these tentatively proposed phonemes seem to take them as firmly established.¹ At least we often see reconstructed forms with these purported phonemes (to quote forms at random that have appeared in recent publications): *xempat, *beHi, *Wasu. It is as if it did not matter what the proto-phonemes were or how they were pronounced — as if the reconstructed form does not reflect a spoken form at all, but is just a shorthand way of listing a set of forms which are similar in various Austronesian languages. But surely a reconstructed form is more than a formula. For whatever reason we may wish to know the history of the Austronesian languages, it is most important to know what the phonemes of the proto-language were and how they must have sounded. This is the only way that we can come to meaningful conclusions as to the subgroupings, types of cultural contact within these languages themselves, contacts with outside groups, and other such questions of prehistory for which historical linguistics is the best evidence. Further, it is the only way that Austronesian historical linguistics can provide meaningful data on the basis of which the science of historical linguistics in general can be developed.

I believe that the basic principle to follow is that only those correspondences which furnish good evidence can be considered to reflect phonemes of the proto-language. It would be tangential at this point to attempt specification of what constitutes good evidence, but it is clear that in the following cases no good evidence has been presented for the establishment of a proto-phoneme:

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 1-30. *Pacific Linguistics*, C-75, 1982.

© John U. Wolff

(1) when the phoneme is posited on the basis of an irregularity of correspondence among reflexes in two or more languages which could easily have been borrowed or readily contaminated by analogical reformations;

(2) when the irregularity of correspondence is found in languages for which the phonological development has not been worked out or where the gloss of the forms in question is not clearly known. This eliminates almost all of the additional phonemes.²

This lays to rest the numerous capitals and subscripted phonemes which Dyen and others have proposed (tentatively or with assurance). With the addition of *N, *C and *Z, we can stick to the phonemes which Dempwolff posited (as corrected by Dyen 1951, 1953a). But we can go further than that in ascertaining the phonology of the proto-language. A number of Dempwolff's correspondences prove to be invalid: to wit, the correspondences symbolised by *d, *r, *c, *z, *g, *T. I showed that the correspondences *d and *r are a chimera in my contribution to the First Austronesian Conference (1974). In this paper I shall give explanations for the other four correspondences *c, *z, *g, and *T. Thus, Proto-Austronesian had the following consonant system:³

p t C k b D Z j s m n ñ ŋ N R l q h w y.

In investigating these four correspondences I start with the same ground rules which I adopted in my 1974 paper. First, forms which were not likely to have been in the proto-language cannot serve as evidence for the existence of a certain phoneme. If there is no evidence for the existence of a certain phoneme except in forms which may have come into the language after the time of the proto-language, there is no evidence for the phoneme. Thus, loanwords can provide no evidence for a proto-phoneme, and loans from outside of Austronesian languages are eliminated from consideration.⁴ Further, forms found only in languages which have had strong influence on each other and which refer to the type of things which likely spread by borrowing cannot be considered to provide evidence. The correspondence must be found in other sorts of forms.

Also forms which refer to flora and fauna, of the sort not of basic importance to the culture, cannot be seriously considered to provide evidence. These terms tend to rapid displacement and to analogical reformation at a rapid rate. This fact can readily be substantiated by elementary fieldwork or by a glance at a handbook such as de Clercq's and the like, which list enormous numbers of synonyms, forms clearly borrowed from language to language, and forms obviously reformed by analogical processes. Also forms which refer to noises or meteorological phenomena tend to be subject to rapid reformation of an onomatopoeic type, so that such forms cannot provide evidence for the existence of proto-phonemes. These considerations eliminate a fairly large portion of the forms presented by Dempwolff for the phonemes *c, *z, *g, and *T. We will now proceed to examine forms containing these forms which have been reconstructed by Dempwolff and the addenda by Blust. My conclusion is that in the case of *c, *z, *g, and *T (just as for *r and *d), the evidence is so weak that we need no longer consider these phonemes to have been part of the inventory of the proto-language.

There are many commonsense considerations which support the correctness of this conclusion. First, the numerous reconstructions which have been attempted from the bottom up have failed to reconstruct any of these phonemes for the proto-languages of subgroups. That is, there is no correspondence in any of the proto-languages of subgroups which reflects PAN *d, *r, *c, *z, *g, or *T — not in proto-Minahassan, proto-Polynesian, Proto-Paiwanic, Proto-Tsouic,

Proto-Manobo, Proto-Bisayan, to quote published studies, nor has anything turned up in the unpublished materials to which I have had access.

Further, all of the languages which Dempwolff chose outside of Melanesia and Polynesia were in contact with and under the influence of Malay, and most of the forms which Dempwolff thought to reflect these proto-phonemes are found in Malay and often can be proven to have been borrowed from Malay into the languages cited. We shall now proceed with an examination of the forms in question.

1. *c

The evidence for *c is very shaky, although a fairly large number of forms were reconstructed with *c. The evidence for *c was thought to be the following correspondence: Ml, Ja c, Ho ts, and the other languages show the same reflex as for *s.⁵ We need not look at all forms which Dempwolff listed with *c: we may eliminate loanwords, and further we need not look at forms which were reconstructed with *nc. Any form reconstructed with *nc can as well be reconstructed with *ns.⁶ Although few of the forms listed with *nc do, in fact, look as though they derive from the proto-language, we need not consider them further here as they provide no evidence for the phoneme *c. Other forms thought to have *c are not good evidence for *c for various reasons:

- (1) they are onomatopoeic or otherwise likely to have been affected by analogical reformations;
- (2) they are found only in languages in close contact with Ml and are likely to be from Ml;
- (3) the meanings of the attested forms are unconnected or the correspondences of the other phonemes in the form are not regular;
- (4) the proto-form could be reconstructed with *s. This leaves us with a small residue of forms which provide the shaky evidence for *c. But these forms are too few and too shaky to be convincing evidence for *c, and I conclude that there was no phoneme *c. Let us look at these forms individually.

1.1. First, we list the onomatopoeic forms and the forms likely to have been affected by analogical changes: *ceguk Tg sigok, Ja ceguk 'hiccough'; *ciap Tg siyap, Ml ciap 'chirp'; *citicit Tg sagitsit 'hissing noise', Ml cicit 'squeak, twitter', NgD kasisit 'hissing noise', No tsetetra 'click of tongue in approval'; *cecak TB sosak (I have been unable to confirm Dempwolff's gloss of 'house lizard'), Ml cecak, NgD tasak 'house lizard', Ho tsatsaka 'kind of lizard'; *licak Ml lècak 'moist and slippery', Ho ditsaka 'wet, soaked, drenched' (Ml has the following forms all of which mean something similar to lècak and which point to analogical reformations of forms with similar meanings: lecap, lincum, lècèk, bècèk, bècak, lècah, lècèt, lèncèt. The chances that Ml lècak has been influenced by some other form of similar meaning are very great.) *pucuk TB putsu 'tip, peak' (=pusuk), Ja pucuk 'point, tip', Ml pucuk 'shoot, top branchlet', NgD tampusok 'end of a fishnet, go upwards to the heights'. (There likely was contamination with Ml puncak 'climax, summit', a point supported by the TB form which reflects a nasal. Further NgD also has kapucok 'a counter for weaponry', a loanword from Ml, and TB pusuk is likely to be a loanword, so that the forms in the various languages with c are all probably from Ml.)

1.2. Now we list forms found only in languages which were under the influence of Ml (or took forms from Ja via Ml). The following list gives those only in Ml and in one other of Dempwolff's languages which was in contact with Ml: *cepat (sic Dempwolff) Ml cepat, Ja cepet, NgD capat 'fast' (Sa toha 'rejoice' is unconnected.) *lucut Ml lucut 'become undone, slip away', NgD lusut 'be loose' (Tg lusot 'pass through', TB lusut 'for the sun to set' are unconnected.) *suci Ja suci, Ml suci 'pure, holy, sacred' (Tg masusiq 'minutious' is borrowed from Ml as attested by the final glottal stop; TB sutsi 'wash' is unconnected [borrowed from Ml cuci 'wash'].) *pucat (sic Dempwolff) Ja pucet, Ml pucat 'pale' (NgD musat 'pale' must be borrowed because of the a of the final syllable.) *ciTak Ja ciTak, Ml cètak 'print'; *canTung Ja canTung, Ml cantung 'kind of dipper'; *canTing Ja canTing, Ml canting 'small dipper for wax'; *cacing Ja, Ml cacing 'worm'. *kucak Ja, Ml kocak 'mix up, stir up', *celak Ja, Ml celak 'eye shadow'; *canTik Ja canTik 'kind of bent knife', NgD nyanyantik 'bent a bit'; *capah Ml capah 'shallow platter', TB sapa 'wooden plate'; *cambuk Ml, Ja cambuk 'large whip', *picek Ml pècak 'one-eyed', Ja picek 'blind' (Tg pisak 'one-eyed' is borrowed from Ml as indicated by the a of the penult.) *ca(n)kup Ml cakup 'catching in the open mouth', NgD mansakop 'snap at' (with irregular morphophonemics), bacakop (=basakop) 'greedy' (Tg sakop 'be under the power of something else', Ja cakup 'get hold of something', Ho tsaku 'be chewed, masticated' do not seem to be connected with the Ml and NgD forms.) The following forms are related by borrowing, for the Tg term refers to a geographical designation, the type of meaning which many loanwords from Ml have: *cabang Tg sabang 'branching of river, bifurcation of roads', Ml cabang 'branch', Ja, Ml cawang 'branching, bifurcation', TB sabang 'lengthwise bind of two rafters'. We include Ho in the languages which were in contact with Ml, as there are forms in Ho clearly borrowed from Ml.⁷ Thus the following forms are connected by borrowing from Ml: *pecut Ja, Ml pecut 'whip', Ho fitsuka 'horsewhip'.

Some of the forms with *c are found in several languages, but they still must be regarded as loanwords because they occur in Ml and only in languages which were in contact with Ml and have the meanings of forms which are likely to be borrowed: *cincin Tg singsing, Ja, Ml cincin, NgD tisin 'finger ring' (The Tg form is also phonologically aberrant.) *cium Ja, Ml cium 'kissing by sniffing, sniff', NgD sium 'kiss', Ho tsiutsiuna 'name of fragrant plant'; *cacar Ml, Ja cacar, NgD kacacar 'smallpox' (stated to be Banjarese by Hardelandt); *cagak Jav cagak 'pillar to hold up a house', Ml cagak 'forked pillar', NgD sagak 'a small protrusion in a cannon which fits into a hole in the support'; *cakep Ja cakep 'can grasp, understand', Ml cakap 'undertake, promise, be able', NgD cakap (=sakap [noted as Banjarese]) 'able, willing'; *candu Ja candu 'essence' Ml candu, TB sandu 'opium', NgD sando 'a fragrant oil'; *cangkul Ml cangkul, Ja cangkol, NgD sangkul 'hole'; *cemeD Ja cemer 'unclean', Ml cemar 'dirt, pollution' (NgD samar 'polluted [by a sickness caused by eating something forbidden]' is borrowed as the vocalisation shows. The forms quoted from the other languages do not have a good semantic connection: TB somor-somor 'bringing bad luck', Fi samo 'stained black', Sm samo 'gum from the eyes'.); *cenTung TB parsottingan 'decorative bow', Ja cenTung 'curly adornment', Ml centung 'erectile tuft or crest of birds'; *ci(m)pi(r) TB sippir, Ml cèpèr 'low rimmed saucer or salver of metal' (also Ml tèpèr), Ja cèpèr (=cepèr) 'saucer shaped'; *cu(ng)kil Old Ja cukil 'dig up', Ja, Ml cungkil 'pry up', NgD sungkil 'log used as lever to lift something' (Ho tsuki 'having a sharp point' does not seem to be connected.) *cukup Ml, Ja cukup, NgD sukup 'sufficient'; *ki(n)cu TB hitsu, Ml kicu 'swindle', Ja kècu 'banditry with force' (? connected); *ka(n)cang TB hacang, Old Ja, Ml, NgD kacang 'bean', NgD kasang (=kacang); *ucap Tg usap 'conversation' TB mangusaphon 'point to good deeds one did for someone', Ja ucap 'what someone says',

Ml ucap 'utterance' (Fi vosa 'speak talk' cannot be connected because the sound correspondences are widely off.)

1.3. Let us now look at the forms which do not have a meaning in common. These provide no evidence for a proto-form, much less for *c. *ca(ng)kem (sic Blust) Ml cengkam 'nipping', Tg sakam 'get with the hand'; *kucup Ml kucup 'kiss' (=kecup), Fi kucuva 'rape a woman'; *li(n)cad Tg linsad 'dislocated, derailed', TB lisat 'pressed flat', Ja lincad 'withdraw', Ml lécat (Dempwolff's gloss 'glassy smooth' not verified.); *lucak Tg lusak 'black mire', TB lusak 'rinse something in water by putting it in a basket and jiggling the basket around', Ho lutsaka (Dempwolff's gloss 'sediment from a solution' not verified); *pecaq TB posa 'grave (of an illness)', Ml pecah 'breaking up into bits', NgD pacah 'shattered' (borrowed from Malay as shown by the vocalism), Ho fetsaka 'thrown down violently, beaten', Fi voca 'be wrecked on a reef', Sa potu-potu 'break by knocking against one another' (Tg pisaq 'crushed' is not connected, as shown by the Cebuano pisaq 'crushed'. Cebuano reflects PAN *e as u.); *pecel Ml pecal 'crush with the fingers, squeeze with the hands', Tg pisil 'squeeze with hands'. (Again Cebuano pisil 'squeeze in hand' shows that the Tg pisil is not connected with Ml pecal.) *cemcem Tg simsim 'eat as a snack, try eating something', Ja cemcem (Dempwolff's gloss 'moisten' unverified). cahing Ja cèng 'thickened sugar cane juice', Tg sahing 'white sticky resin' is from saleng. NgD saing 'taut' is not connected.

1.4. In the following forms the reflexes which were thought to prove *c are unconnected and the reconstruction may as well be with *s (if indeed any of the attested forms reflect a proto-form): *cabat Tg sabat 'intercept', TB sabat 'hindrance'. (The following forms are not semantically connected with the above two: Ja cawad 'omission', Ml cawat 'loincloth' [=Jv cawet 'loincloth'], Fi cambaka 'blurt out a secret', Fi saambaya 'ward off a blow'.) *cacah Tg salasaq 'destruction', TB sasa 'be fallen in' (Ja cacah 'cut up', Ho mitsatsa divided into two or more parts are unconnected with the above and probably unconnected with each other.) *calcal Tg salsal, NgD tasal 'forge' (Ja cacal 'chipped, hacked, be planed off from' is unconnected.) *cencen Tg sinsin 'closeness of intervals', TB sotson 'be situated close to one' (Ho tsentsina 'plugged, stopped up' and Fi sosoo ucu 'put finger in nose to keep out bad smell' are unconnected.)

1.5. We now come to the forms which cannot be dismissed out of hand. There are only two forms which have correspondences outside of Indonesian and are good strong cases: *cabut TB sarbut, Ja cabut, Ml cabut 'extract, pull out by force', Fi tacavu 'uprooted', cavuta 'extraction by force'. This is a convincing form, but nevertheless I think that the initial c's of the Ml and Ja are secondary. The chances that analogical reformation took place in the first syllable are very good in view of the large number of forms which have a meaning of 'uproot' which end in a reflex of *but or *bet: Ja jabut 'uproot', Ml rebut 'grab', Cebuano gábut, ibut 'uproot', to mention a random sample. The other form is *camuk Tg samuk (Noceda's gloss: 'a large mouthful'), Ja camuk 'chew on', Ho tsamutsamuka 'smack the lips', Fi namuta 'chew and spit out again', Ulawa dremu 'eat areca nut', To hamu, Sm samu 'eat vegetables or yams alone'. In this case the meanings of the various forms are quite far apart and the resemblances are probably

fortuitous (at least between the Oceanic forms and the Indonesian forms.) In the following case the Tg form shows *l* corresponding to Ja *D*, something not elsewhere found in borrowed forms, but which is the normal correspondence in inherited forms. However, the Tg meaning is not close to the Ja and Ml and the resemblance is probably fortuitous: *ci(n)Duk Tg sílok 'spoon', Ja ciDuk 'ladle, scoop, dipper', Ml ciduk 'digging up with a spade'. (Ho tsindruga 'little things which are picked up for eating' is not connected.) In the following two cases the meaning is of a sort which does not point to a borrowing. However, the likelihood that the first syllable with a *c* in Ml and Ja is a reformation is great because of the synonyms or words of similar meaning which share the second syllable: *cahuR Tg sahog 'mixture' TB saor 'mixed', NgD caur (=saur) (said to be from Banjarese) 'mixed' (cf. Tg bahug 'mixed, diluted', Ml baur 'mix', Cebuano láhug, káhug 'mix'.) Similarly, *cuping TB suping, Ml cuping 'earlobe', Ho sufina 'ear' (Sa suhi 'shave the head' is unconnected.) Note that the Ho has a reflex with *s* which is evidence that the *c* of Ml cuping is secondary. (Ja has kuping 'ear'.) In the following case the Tg form only vaguely resembles the Ml and Ja in meaning and there may well be no connection: *campur Tg sapol (Laktaw: 'inclusive') TB suppur, Ml, Ja campur, NgD campur (=sampur) 'mixed'. Further, the Tg final *l* fits into no well-established correspondence with the other forms.

2. *z

Dyen (1951) proposed a proto-phoneme *Z to account for the correspondence Ml, NgD *j*, Ja, TB *d*, Tg *d*, *l*, Ho *r* as opposed to *z which would account for Dempwolff's correspondence Ml, NgD, Tb, Ja *j*, Tg *d*, *r* Ho *z*. Dyen thought that *nZ is reflected as Ho ndr and, *NZ as Ho nj. As in the case of *c, a serious question arises as to whether the forms in the correspondence *z are actually related by inheritance from the proto-language. Again we may note that the forms in the *z correspondence have no good cognates outside of those languages which are all strongly influenced by Ml. Further, the semantic character of the forms which exhibit the correspondence for *z is often such that the forms are likely to be borrowed from Ml. Some of the other forms cited for the *z correspondence in fact enter the correspondence for *Z, and a number of the forms in the *z correspondence are unconnected in meaning.

2.1. First, we will look at the forms which are very likely to be borrowings. Those found only in Ml and one other language are the following forms: *ta(n)zak Ja tanjak 'holding oneself upright', Ml tanjak 'sticking up, rising up to a point' (if these are connected); *tinzu TB, Ml tinju 'box'; *z(a,e)(m)but TB jabut 'hair on chest', TB jambut 'hair of animals', Ja jembut 'pubic hair' (There is no Ml form attested in this correspondence, but the vocalism of TB proves it to be a borrowing and probably it was from a Ml dialect with a for e.) *zalu Ml, Ja jalu 'cock's spur'; *zangkah Ja jangkah 'step, stride', Ml jangkah 'step over with difficulty' (The Polynesian forms which mean 'dance' are unconnected: To haka, Fu saka, Sm sa'a.) *zantung Ml jantung 'heart', Ja jantung 'heart, bud of banana blossom'; *zatuh Ml jatuh, NgD jato 'fall' (The o of the final syllable in NgD points to borrowing.) *zempuT TB jopput 'pick up with two or three fingers', Ml jempuT 'pinch between the fingers, press hand of visitors' (Tg dampuT 'pick up with the hand' is not connected; neither is Fi covuta 'break or cut food small, peck at'.) *zemuR TB jomur 'dry rice in the sun', Ml jemur 'dry by exposure to air'; *zulzul Ja jujul 'sticking out by being too large and not

fitting', Ml jujul, jojol '*sticking out in a pointed way*' (Tg duldul '*shove with force*' and TB juljul '*add to something forcefully*' are not connected.) *zu(m)put Ja jumput '*pick up something small with fingers or tweezers*', NgD jumput '*pick up something small*' (=sumput) (There is no Ml form attested for this correspondence, but the doublet in NgD attests outside influence. Further, there are doublets with similar meaning e.g. Ja jumput, jukuk, jupuk '*pick up*' and also the forms listed under *zempt above, so that the initial j is the result of analogical developments in all likelihood.) *zungzung TB, Ml jungjung '*support on head*' (Ho junjuna '*haughtiness, pride*' is not connected.) Other forms are widespread in Indonesian languages, but they have the meanings of forms which are subject to learned borrowing: *azi Tg pangadyiq '*pray*' (with irregular correspondences, attesting to the fact that it is a loanword), Old Ja aji '*sacred book*', TB aji '*magical formula*', Ml aji (said to be a Ja loanword), kaji '*study religious works*', NgD kaji '*learn, study*' (Fi kaci '*call, invite*' is not connected.) *bazi TB, Ml, NgD baji '*wedge*'; *kazang Tg kárang, TB hajang, Ja, Ml, NgD kajang '*awning on a small boat*'; *panzi Tg pandi-pandi '*banneret*' TB panji-panji '*ornament of feathers like that worn by children on their heads*', Ja panji '*flag*', Ml panji-panji '*streamer, pennant*'; *puzi⁸ Tg puri, TB, Ml puji '*honour, fame, praise*', Ja puji '*prayer, wish, hope*'; *suzi Tg suri '*needle used to separate the threads in weaving*', Old Ja suji '*needle, thorn*', Ml bersuji '*embroidered*'; *tanzung TB pulo tanjung '*peninsula*' Ml tanjung, NgD tanjong, Ho tanjuna '*cape*'; *tazak Tg tárak '*stabbing forcefully with knife*' (probably not connected), TB tajak '*knife for hacking*', Ml tajak '*grass cutter*', NgD tajak '*strike*', (Sa ata-ata is not connected. I was unable to confirm Dempwolff's gloss of '*grater for yams*'.); *tazi Tg táriq, TB, Ja, Ml, NgD taji '*blade used in cockfighting*'; *uzi Tg úriq '*kind, class, karat*', TB manguji, Ml, Ja uji '*test*', NgD puji '*admonish*'; *zabi TB jabi-jabi, Ho zavi, Fi savirewa '*kind of ficus tree*' (These are names of plants not basic to the culture and thus no evidence for a proto-phoneme.) *zalang TB jalang '*running wild (as of boys that go away from home)*'; Ja, Ml jalang '*running wild (especially of prostitutes)*'; *zagal TB jagal '*meat sold in trade*', Old Ja jagal '*slaughter*', Ml jagal '*deal in trade*'; *zambay Old Ja jambé '*areca palm*', Ml jambi '*areca nut*'; *zamban TB, Ja, Ml jamban '*privy (over a stream)*', NgD jamban '*hut over a river*' (Sa tapa-tapa is not connected. I have not been able to confirm Dempwolff's gloss '*fenced-in place for offal at death feasts*'.); *zanggut TB, Ml, Ja, NgD janggut '*beard*', zangka TB jakka, Ml jangka '*measure off*', Ja, NgD jangka '*compass or callipers for measuring off*'; *zava Tg dáwa, NgD jawe '*millet*' (No Ml form survives, but this is a Sanskrit loan, although Dempwolff does not list it as such.); *zinten, MgD, Ml jintan, Ja jinten '*caraway*'; *zelai Ml jelai, Ja jali, NgD jilai '*kind of grass*' (This form has all sorts of unusual sound correspondences, as is common with plant names which travel from language to language.) *zengkal Tg dangkal, TB jokkal, Ml jengkal, Ho zehi '*span*'; *zu(ng)kung Ja jukung, jungkung, Ml jongkong, jukung, jungkang, etc., NgD jukung '*dugout canoe*'; *uzung TB ujung '*end, exit*', Ja, Ml ujung '*end, tip*' (Fi ucu '*nose*' is not connected.)

2.2. The resemblances among some of the forms which Dempwolff cites are purely fortuitous, and the meanings are not connected: *zagzag Old Ja jagjag '*touch ground with feet*', TB jakjak '*for flowers or rice panicles to be erect*', NgD jajak '*go, tread on*' (from the language of magic)', zamzam Tg damdam '*feel pain*', TB jamjam '*sharp, hot*'; *zawazw Ja jojo '*trying to resell something*', NgD jao-jaoi '*for the chin to move, rechew*'; *zegzeg Ja jejeg '*straight, upright, always, steady*', Ml jejak '*step, footprint*'; *zehet TB jehet '*earth that is sticky and*

clay-like and hard to work', Fi koko 'join, fasten', PPN *soko 'tie, fasten'; *zengat Ja njengat 'in an upright position', Fi cenga 'lift things up by one side only (as the leaves of a book, or a lid)'; *zelag Tg dilag 'beauty', Ho jelaka 'glaringly bright'; *zeng Ja jeng 'foot', TB jongjong 'stand upright'; *zingzing TB jingjing 'held in the air, carry held in the air', Ja jinjing 'walk on toes' Ml jinjing, jènjèng 'carry a light burden dangling from the hand'; *zu(ng)kit Tg dukit 'hole made by poking the finger into something', TB jukkit 'pull fruits down from tree'; *zu(ng)kuk Ja jungkruk 'with bent head', Ml jongkok 'squat', Ho juku 'approach quietly, stooped down'.

2.3. For a number of the forms reconstructed with *z, the TB, Ja, or Ho forms are borrowed from Ml, are unconnected, or affected by analogical developments. If the other forms cited are cognate, the proto-form can be reconstructed with *Z. *hizaw Tg hilaw 'unripe', Ml hijau 'green'. The following forms are borrowings from Ml: Ja ijo, NgD hijau 'green'. (Ho maitsu 'green' does not fit into this correspondence.) *tin(z)ak Cebuano tindak 'tread on', Minangkabau tijak 'step, tread', NgD (said to be from Banjarese) tinjak 'stepped on', Ja tindak 'go (honorific)'. (Ho tsinjaka 'a dance with stamping of feet' could well be a borrowing from Ml.) *zahat Ml jahat, Fi caa, Cebuano daqut 'bad'. A reconstruction *Zaqet would account for these forms. The Polynesian forms which Dempwolff cites are not directly comparable, neither phonologically nor in meaning. TB jat 'bad' and Ja jahat 'grave (illness)' must be borrowed from Ml because of the vocalism.) *zahit Tg dait 'close, contiguous', TB, Ja, Ml jait 'sew', Ho zaitra 'sewn', Fi caita 'coitus', PPN *saqi 'bind'. We may reconstruct *Zaqit if we take the TB, Ja and Ho forms to be borrowed from Ml. Since all these forms show a semantic development in common, they are most probably indeed borrowings. *zalateng This is not an inherited form, but if it were, there is no reason not to reconstruct with *Z: M jelatang, Fi salato, Sm salato 'kind of tree'; *zalzal Tg daldal 'blow', To haha, Sm sasa 'beat, thrash'. (If these are connected at all, reconstruct *ZalZal. TB jaljal 'cut meat up fine' is unconnected.) *zaRum Ja dom, Ml jarum, Fi sau, Sa sulitelu, To hau 'needle'. Tg karayom 'needle' has an irregular correspondence for *R and has a petrified prefix. It is not directly inherited. Similarly, TB jarum is a borrowing from Ml. This form should be reconstructed as *ZaRum. *zual Ja dol, Ml, NgD jual 'sell'. (TB jual 'measurement for rice' is certainly not connected by inheritance, if it is connected at all.) We may reconstruct *Zual, if this form is inherited at all. *zuling Ml juling, Tg duling 'cross-eyed'. There is no reason not to reconstruct *Zuling, if the forms are inherited. *enzut Ml enjut 'move with a jerk', Tg undot, kundot 'fall back in fear'. There is no reason not to reconstruct *enZut if these forms are connected. *panzang Ml panjang, anjang, NgD panjang 'long', Ho andrandrana 'erectness, uprightness of men'. The Ja panjang 'long' is of the Krama (high) speech level and likely to be borrowed from Ml. Also Ja shows the accretion of initial p as does Ml. We may reconstruct *anZang. *tinzaw TB tindo 'look in the distance'; Ml tinjau 'look', Fi tirova 'look at in mirror', Sa 'iro 'look at, for'. Reconstruct *tinZaw. Ja tinjo 'visit' is an honorific and borrowed. Ho tsinju 'gaze at from distance' has the reflex of a borrowing. The TB form with nd proves that the proto-form had *nZ, and the Ho must be a loanword.

2.4. The following forms do not have the semantic character of learned borrowings, and they cannot be dismissed outright. However, they are confined to languages which are in contact with Ml and even if they are of the basic vocabulary, they probably are borrowed nevertheless. In a few cases there is other evidence that these forms, which point to *z, are, in fact, not inherited, but borrowed. There are only two forms in this group which are not readily explained as loanwords. They are *tazem 'sharp' and *za(m)bay 'for branches to hang down'. But it is certainly not impossible that they are loanwords or affected by analogical changes.⁹ These are the forms in this group: *azar Ml, Ja, TB, NgD ajar, Tg áral 'teach, learn'. The Sa manata'ini 'know' is not connected; *banzar TB, Ja, Ml banjar 'row'. (Fi basa 'level, in a line with' is probably not connected.) *ta(n)zuk Ml tajuk, Ja tanjuk, NgD tajok 'tuft'. (The o of the final syllable in NgD is indicative of a loanword.) *tazem Tg talim, TB tajom, Ja tajem, Ml tajam, NgD tajim 'sharp'. Fi tasi 'shave, scrape wood' is unconnected. *za(m)bay TB jambe, NgD jawei 'for branches to be hanging down'. (See n.9 above for a discussion of this form.) *zanzi TB, Ja, Ml, NgD janji 'promise'; *Zalin TB, Ml, NgD jalin 'bind, braid'; *za(m)bat Ml jabat, jambat 'take in hands, hold on to', PPN *sapai 'lift, carry'. If these forms are related, the reconstructed form can be *Zambat. TB jabatan, Jav jawatan 'office, duty' are borrowed from a Ml form with a derivative meaning jabatan 'office, duty'. *zamah Tg dama 'perceived, felt', TB jama 'touch', Ja jamah 'having relations with a woman not one's wife', Ml jamah 'physical possession, handling of a woman not one's wife' (The meaning of the Ja form proves that it is borrowed. The Tg must be borrowed because of the lack of final -q, and thus it is likely that the TB is also a loanword.) *azak Ml, NgD, ajak Ja jak 'urge, invite, incite'; *zayu Tg dayo (Cebuano dayqu), TB jau 'foreigner', Tg láyoq 'far', Sm sau 'ai tangata 'cannibal (eating foreigners)'. The TB form is probably not connected, but a borrowing from Ml jauh 'far'. If the Sm and Philippine forms are related, we could reconstruct *Zayqu or *Zayuq. *zinak TB, Ja, Ml, NgD jinak 'tame'. *zuluk TB julluk 'bore, force into (like worms)', Ja julukaké 'push something upwards', Ml jolok 'poling, thrusting upwards, forwards'. The other forms cited in this group are not connected in meaning: Tg dulok 'remove, excavate', NgD julok 'be reached, brought here', Ho juluka 'enter, go in', Sm sulu 'tuck in, insert, sheathe'.

The forms cited in this section are the only ones that cannot be dismissed out of hand, but they do not offer strong enough evidence to allow us to reconstruct a proto-phoneme *z. This phoneme did not exist in the proto-language.

3. *g

The evidence for *g is better than for *c and *z, but still not strong enough to warrant the establishment of a proto-phoneme *g. First, correspondences of forms with *g are not attested outside of Indonesia-Philippines. A good number of the forms with *g are borrowed from Chinese or Indic sources. Others can be proven to be loanwords on internal grounds (other than that provided by *g). Further, there is evidence that the reflexes of *g had a marginal status, although everywhere they are well established now. Ml and Ja have many forms with doublets showing g and b, g and d, or g and k.¹⁰ Also, many forms with g in Ja and Ml have correspondences with k in other languages, and occasionally, vice versa, forms with g in other languages have k in Ja or Ml. All of this mitigates against the likelihood that there was a proto-phoneme *g. We cannot dismiss out of hand all of the forms in the *g correspondence, and we will present those forms which cannot be readily explained below in Section 3.5. We need not, however, consider forms reconstructed with *ngg. Since there are no forms

which were reconstructed with *nR, and only one form which has tentatively been reconstructed with *nj (*anjaw 'day'), the forms with *ngg may actually reflect *nR or possibly *nj (if they are indeed inherited forms). In any case, they offer no evidence for a phoneme *g.

3.1. A great number of the forms in the *g correspondence are those which are likely to have been affected by analogic changes. The following forms are names of sounds: *gagak TB gagak 'laugh loudly', Ja, Ml gagak, Ho (dialectal) gaga 'crow', NgD gagak 'cackling of chickens', Sm 'a'a, To kaka 'chattering of birds', *gukguk Tg gukguk 'grunt of pigs', TB gukguk 'kind of bird sound', Ja nggukguk 'uncontrolled weeping', Sm 'u'u, Fu kuku 'kind of bird'; *legung Ml legung 'boom, gong', Tg lagong 'bass, low sound of voice' (The correspondence Tg a, Ml e shows that the similarity of form in these two words is coincidental.) *gembar Ml gembar-gembar 'yelling noise', Tg *gimbal 'clatter'. I believe that words meaning 'shake, shine, tremble', and the like are subject to similar kinds of analogic reformations: *gigil 'shiver', Tg gigil 'tremble with pleasure or delight'; gilang Ja, Ml gilang 'bright, glittering', To, Fu kikila, Sm 'i'ila 'shine, glisten'; *guyang Ja goyang 'unsteady, shaking', Ml goyang 'oscillation' (NgD gagoyang 'often go somewhere' is probably unconnected.) The following forms also are probably affected by analogic reformation: *gumi Tg gumi 'whiskers (as of a cat)', NgD gumi 'beard, whiskers (of cat)'. These forms may well be contaminations of reflexes of *kumis: Ml kumis, Tb gumis 'mustache, beard', Fi kumi 'chin, beard'. *(Ct)eguk Ml teguk 'gulp down', Tg tagok 'gulp, swallow', Tg tigok 'sound produced by swallowing' (In this case the numerous forms with a similar sound and similar meaning prove that these forms are analogical creations, e.g. Ml ragok, degok, cegok, Tg lagok 'gulping, swallowing, and the like'. cf. *ceguk Section 1.1., above.) Another example of numerous doublets which point to analogical contamination: *pegang Ml pegang 'grasp, hold', NgD hegang 'hold', TB rogang 'hold fast', NgD pagang, pegang 'under the power of' (borrowed from Ml). The following forms end in -tas, a sequence which many forms of similar meaning end in: *ge(n)tas Ja getas 'break apart easily', Ml getas 'brittle', Ml gentas 'snapping a twig', NgD getas 'be broken through', Ho hentana 'crack, fissure'. cf. Ml retas 'rip', Ja tuntas 'through and through', Cebuano bugtas 'snap', bitas 'lacerate', etc.) Thus it is likely that the initial syllable is a reformation. The following forms also appear to be secondary: *galing, *giling, *guling, and *gulung. The similarities of meaning point to analogical reformation and similar words have a k: e.g. TB haling 'turned', NgD kuling 'be rolled, wrapped up', Jav kilingan 'turning round and round', Ml keliling 'going round'. *galing TB haling 'turned' NgD galing 'roll', Sa 'ali 'be lying curled up'; *giling TB giling 'roll, turn around', NgD giling 'roll up', Ml giling 'rolling out, flattening with roller', Ml, Tg, Ja giling 'grind' (The Ja and Tg are probably borrowed from Ml.) *guling TB guling 'bend, roll to the side', Ml guling 'roll over', Ho hudina 'turning around'. (Fi qili 'rub' and qilica 'twist fibres in hands' is unconnected with these forms.) (The following forms are not connected with these in any way: Old Ja guling 'lie down', Ja guling, NgD gaguling 'pillow to put one's leg over while lying down'.)

3.1.1. There is a vacillation between g and k in many forms. The forms with k are original and the forms with g are secondary. If these forms are reflexes of proto-forms, they should be reconstructed with *k-: *gatel Tg kati, TB gatal, Ja gatel, NgD gatel, Ho hatina 'itch'. (The vocalism of the TB in this case confirms that the TB form is a loanword.); *ga(m)pit TB gappit 'clamped', Ja gapit 'tongs'. These forms must be reflexes of *ka(m)pit: Tg kápit 'hold, grasp', TB hapit 'clamp together', Ml kámpit 'sack' (maybe not connected), Fi kabita 'adhere to', Sm 'apili 'cling to, stick to'. (The To and Fu forms which Dempwolff cites are not easily connectible in meaning, but the Sm form is representative of cognates in a large number of Polynesian languages.) *galang/kalang TB halang, NgD galang 'something on which something else is set', Ml kalang 'support, prop', Ml galang 'thwarting beam', Ho halana 'roll on head to carry burdens'. In the case of *gapit and *galang, it is quite certain that the correct reconstruction is *kapit and *kalang respectively because their roots contain a petrified prefix *ka-. Most of these forms are of the sort that are likely to be loanwords or liable to analogical reformation. The following forms are likely to be borrowed: *gasing Cebuano kasing, TB, Ml gasing, Ja gasingan 'top' (Ho hasina 'be very round' may well not be connected.) *genDang TB gondang 'Batak orchestra music', Ja kenDang, Ml gendang, NgD gandang 'kind of drum'; *genDit, TB gondit 'kind of cloth belt', Ja kenDit 'sash, cloth belt', Ml gendit, kendit 'girdle'. Forms likely to have been subject to analogical reformation: *geli Ja geli 'have a compulsion to laugh', Ml geli 'tickled, amused', NgD kalien 'ticklish' (Sm 'oli'oli 'rejoice' is not connected.) *gitik Cebuano gitik 'tickle', TB gitik 'tease', Ml gelitik 'tickle', NgD kitik 'tickle'; *gilap Ja gilap, Ml gilap 'shining and gleaming', NgD gilap 'whetted, highly sharpened', Tg, Ml kilap, TB hilap 'flashing, flaming'; *gusok TB, Ja, Ml gosok 'clean by rubbing', NgD kusok, Ho kusuka 'rubbing'; *guncang TB gutsang, Ml guncang, goncang, kuncang, koncang 'shake' (Ho huntsana 'rinsed' is probably unconnected.)

3.2. A large number of forms in the *g correspondence have no semantic connection and are not cognate. These forms did not exist in the proto-language: *gaDgaD Tg gadgad 'shelled out (of grains)', TB gargar 'fallen apart into pieces', Jav gagar 'fail to complete a normal cycle', NgD gagar 'fall out of, fall off'; *gahul Tg gahol 'lacking in time', Jav kagol 'hindered'; *galgal TB galgal 'struggle against', Ml, Ja gagal 'fail'; *gasang TB gasang 'easily catching fire', Ja geseng 'burned, scorched', Ml gasang, gangsang 'libidinous', Fi qesa 'charred', To kakaha 'glow with heat, flare up', Sm 'a'asa 'red hot'; *gas TG gasgas 'worn out by friction', TB gas 'broken in two'; Ja gas 'make afraid', NgD gagas 'wood from interior swamps'; *gelgel Tg gilgil 'bearing down with hand' Laktaw: 'cut off', Ja gegel 'notch', To koko 'squeeze, press', Sm 'o'omi 'squeeze, crush'. *gemgem Tg galimgim 'loneliness, homesickness', Ja glenggem 'being quiet so as to steal others' property', NgD gagom 'be dumb, not know what to say'; *gesa Tg gisa 'itching, stinging in body', TB gosa 'torture', Ml gesa 'hasten', Fi qosa 'eat noisily'; *gigi TB gigi 'ends of wings', Ml gigi 'tooth', Ho hihi 'scrape, scratch'; *gilang TB gilang 'be undecided, go here and there in indecision', Ho hilana 'removed to one side'; *gilis Ja gilis, giles 'mash, flatten', Fi kilice 'exhume'; *guDang Tg gulang 'old', TB godang 'much, be grown up' Ml gedang 'large', Ho hurana 'become augmented (as a fire, plague, etc.)' (There is some similarity of meaning here, but I believe the similarity is fortuitous.) *gugut Tg kukot 'eat little by little for amusement', TB gugut 'gnaw at', Ja gugut 'kill lice by pressing against the front teeth', NgD gogot 'dried up, used up (of the soil)'; *gutem Tg gutom 'hungry', TB gutom 'scavenge', NgD getem 'to

harvest'; *gusgus Tg gusgus '*rinse the hair with gogo*', TB gusgus '*scrub, rub*', Ja gugus '*corroded*', gogos '*eroded*'; *sa(n)gap Tg sagap '*scooped up from just below the surface*', Ho tsaka '*be fetched from a well (water)*'; *teges TB togos '*dedicate to*', Ja teges '*meaning*', Fi doka '*respect, honour*'; *ulug Ml olok '*tease*' Tg úlok '*coax, cajole, induce*', Cebuano ulug-úlug '*flatter*'; *DegDeg Tg liglig '*shake, jolt, rinse*', Ja DegDeg '*thumping heart beats*'; *cekig Ml cekik '*strangle*', Tg sikig '*tight-fitting (around neck or armpits)*', Ja cekik '*hiccough*'; *sugsug Ml susok '*stab, pierce*', Tg salugsug '*sliver*', Ja susuk '*something stuck under the skin as a charm*'; *tibag Tg tibag '*demolished*', Ml tebak '*crowbar used to break stone*'.

3.3. A great number of the forms with the correspondence *g are clearly borrowed from Ml. They occur only in languages which were in contact with Ml and have meanings typical of loanwords: *agak Ml agak '*guess*', Ja agak '*thought, opinion*', TB agak '*estimate*'; *bagus Ml, Ja bagus '*handsome*'; *ager TB, Ml, Ja agar-agar, Ja ager-ager '*agar-agar*'; *gayang TB geang-geang, Ml gayang '*kind of hanging shelf or basket*' (NgD gayang '*kind of sickness*' is not connected.) *gaduk TB gaduk '*haughty, conceited*', Ml gaduk '*swagger*'; *gaDay TB gade, Ja gaDe, Ml, NgD gadai '*pawn*'; *gaDung TB gadong, Ja gaDung, Ml gadung '*kind of yam*', NgD gadung '*kind of medicinal plant*'; *gagah Tg gaga '*usurpation*', TB gaga barani '*brave*', Ja gagah '*strong, muscular*', Ml gagah '*great physical strength*', NgD gagah '*stand up and fight, not give in*'; *ganzil Tg garil '*defective in pronunciation*', Ml, Ja, TB ganjil '*uneven, odd*', NgD ganjil '*not easily divisible, lacking*'; *galumbang TB galumbang Ml gelumbang NgD galombang '*long waves*'; *galumat Ml gelumat, NgD gaiomat '*decking*'; *gana TB gana '*a picture one swears by*', Ja gana '*figurine*', NgD gana '*the spirit (soul) of a thing in human form which leaves the thing and appears in dreams*'; *gantang TB gattang, Ja, Ml, NgD gantang '*a dry measure*', Cebuano ganta '*a dry measure*'; *gangsai Tg gansai '*excess, unpaired*', Ml gasal '*uneven*', Old Ja gangsal '*five*'; *gawai Tg gawai '*wizardry, witchery*', Ja gawe '*work*', Ml gawai '*work*' (from Ja), NgD gawayan, goayan '*rice or work given in payment for interest on borrowed money*'; *geber Ja geber '*screen, curtain*', Ml gebar '*coverlet, tapestry*', NgD gabar '*kind of coarse material*'; *galang Tg galang '*bracelet*', TB golang, Ja gelang '*armband*', Ml gelang '*circlet, circuit, bracelet*', NgD gelang '*copper leg rings*'; *genep Tg ganap '*completed, fulfilled*', TB gonop, ganup '*each and every*', Ja genep, ganep '*completed, fulfilled*', Ml genap '*completing, even number*', NgD genep '*each and every*', kagenep '*even number*', Ho henika '*full, complete*' (The irregular correspondences in Tg, Ja, and TB confirm that this form spread by borrowing.) *getaq Tg gataq '*coconut milk*' (with the correspondence of a for Ml e indicating a loanword), TB gota '*sap*', Ml getah '*sap, latex*', NgD gita (also not a normal reflex of *e) '*sticky tree sap*'; *gunung Ja, Ml gunung '*mountain*', NgD gunong '*name of mountainous geographical area near Banjarmasin*'; *gila TB gila '*excited out of joy*', Ja gila '*be crazy for something*', Ml, NgD gila '*crazy*', Fi kiila '*wild, savage (of animals)*'; *gusi TB gosi (I was unable to confirm Dempwolff's gloss '*gums*'). Ja, Ml gusi '*gums*'; *gusa(r) TB, Ja, Ml gusar '*furious*' (Fi kusa '*be quick, hasten*' is unconnected.) *gunDul Ja *gunDol, Ml gundul, gondol, NgD gundul '*shorn, bare*'; *gunDik Ml, NgD gundik, Ja gunDik '*concubine*'; *gulai Tg gulai '*vegetables*', TB gule '*sauce for vegetables*', Ml gulai '*wet curry*', NgD gulai '*stirred*'; *gunting Tg, Ja, Ml gunting, TB gutting, gusting '*scissors*', NgD gunting, kunting '*be cut*'; *guRi TB guri-guri, Ml guri, kuri '*small earthenware vessel*'; *guluk TG gulok '*short machete*', TB goluk (with o for u, an indication of a borrowing) '*a knife to hack with*', Ja golok '*machete, dagger*',

Ml golok 'sword with convex cutting edge, kind of machete'; *guham Tg guham 'fiery blisters on the body', TB, Ml guam, Ja gom, NgD goam 'thrush, eruption in mouth' (Tg guham is perhaps directly related to Ml ruam 'rash'); *suligi Tg sulígiq (with a final -q, an indication of borrowing) 'dart, lance', TB, Ja suligi 'bamboo lance', Ml seligi 'dart, javelin'; *sagu Ml sagu 'mealy pith (usually of sago palm)', Ja, TB, Tg sagu, NgD sago 'sago', Ho saku 'corn'. The following forms are probably related by borrowing because the vowel of the first syllable fluctuates between e and a: *ganti/genti Tg ganti 'reciprocal act', TB gotti 'one who replaces', Ja ganti, genti, Ml, NgD (noted as Banjarese) ganti 'replace', *sigai Tg sigay, Ml sigau 'small cowrie'. The following set comes from a reformation of *belaj 'spread out' which has passed into other languages from Ja via Ml as the meanings attest: *gelar Ja gelar 'spread out, title',¹¹ TB goar, Ml gelar, NgD galar, garar 'title, name by which one is honoured'. The following form is no longer attested in Ml except in titles, but the final -q in Tg proves that it is a loanword: *gawa Tg gawaq 'do', TB margao 'work (used in idioms)', Ml panggawa 'a military rank', NgD panggawa 'chief (in the language of magic)' (Ho kava-kava 'move the hands and feet about, as soldiers out of step' is not connected.) Forms with reflexes found only in Ja and Ml or these two languages and NgD cannot be used as evidence for *g. We place them here: *gigit Ja, Ml gigit 'bite'; *gili(r) Ja, Ml gilir, NgD girir 'be in turns, cycles'; *gayung Ml, Ja gayung, NgD gayong 'dipper'. In the following two forms the TB is not connected in meaning or is a borrowing and the reflex is attested only for Ml and Ja. We therefore place them here: *gemuk Ml gemuk 'fat, fertile', Ja gemuk 'lubricating grease', TB gamuk 'fat' (a borrowing as the vocalism attests); *tegeng Ja tegent 'strong, firm', Ml tegang 'taut, out-stretched' (TB togong 'hard to digest' is not connected in meaning.) The following form is not attested for Ml, but the meaning is such that borrowing is highly likely: *gantal TB gatal 'food given to gamblers', Ja gantal 'rolled up betel leaf' (The Polynesian forms are not connected: To kata 'branch from the kava plant', Fu kata, Sm 'ata 'cutting from the kava plant'.)

3.4. Two forms reconstructed with final *-g may as well be reconstructed with final *-R, if they reflect a form from the proto-language at all: *DugDug Tg luglug 'move to and fro', To lulu-lulu 'shake' (Ja DoDog 'knocking sound' is not connected.) *segseg Tg sigsig 'torch made of bundles of reeds or bamboo cylinders', Ho sesitra 'smoke hemp' (TB masoksok 'burned up' is not connected.)

3.5. This leaves us with fifteen forms which were reconstructed with *g and which do not have reflexes in Ml, or if they have Ml reflexes, the meaning is not one which would lead one to suspect a loanword. None of these forms are of the basic vocabulary, however. They are the only evidence for *g which cannot be dismissed out of hand, but since none of these forms show cognates in Oceania or Formosa, I believe that they are likely to be related by borrowing, and I do not consider them sufficient evidence to reconstruct *g. In some cases the coincidence of meaning is probably fortuitous. *agam Tg agam-ágam 'retrospection', TB agam 'have an opinion'; *ga(r)ang TB garang 'easily inflammable (like powder)', Ja garang 'dry over fire' (These two forms may resemble each other coincidentally. The meaning connection is not good. Ml garang 'fierce, turbulent' and NgD gagarang 'for the noon to be very bright' are not connected.) *galaq TB gala, Ml galah 'long pole'; *gangu TB ganggu 'be in doubt, not trust', Ja, Ml ganggu

'bother', NgD ganggo 'hindered' (The vocalism of the NgD indicates a borrowing, and the TM may be unrelated.) *galak TB, NgD galak 'burn brightly', Ml, Ja galak 'untamed, dangerous, meracing'. (Tg galak 'joy' is not connected with the TB and NgD forms.) *ganggang TB gang, ganggang 'being apart, being too loose', Ja ganggang 'parted (as of people that are fighting)'; *gapay TB, Ja gape 'slack, weak' (These forms may be analogical reformations as many forms with a similar meaning end in -pay: Cebuano tupay 'unconscious', Ml sampai 'hang loose', Ml melampai-lampai 'swaying'.) *gantung TB gattung, Ja, Ml gantung, NgD gantong (with o of the final syllable pointing to a loanword), Ho hantuna 'hand, suspended'; *gemgem TB gomgom, Old Ja gegem 'control, rule', Ja gegem 'control, hold in fist', Ml genggam 'hold in the fist', NgD sagengem 'what one can enclose with the four fingers of each hand'. (In this case the TB and NgD forms are clearly borrowings as their meanings attest. The TB shows a semantic development in common with Old Ja, and the NgD is a term of measurement. The forms from Oceania are not connected in meaning: Fi qoma (I was unable to verify Dempwolff's gloss of 'grasp'), To kokomi, Sm 'o'omi 'squeeze, crush'.) *gelap TB golap, Ml gelap 'darkness', Ja gelap 'undercover, illicit' (The Ja has a developed meaning which makes one strongly suspect borrowing.) *genting Tg ginting 'an uneven thread', TB gotting 'small in the middle (as if pushed in by a girdle)', Ja genting 'frayed, broken nearly through', Ml genting 'slender link between two larger objects', NgD ginteng 'thin in the middle' (with an irregular vowel in the first syllable); *gayung TB gaung 'hang swaying', NgD gagayong 'hang free in the air from something' (with an o in the final syllable indicating not a direct inheritance); *hegab TB ogap, mogap, 'choke, drown', Ja megap 'pant, breathe in gasps' (Tg higab 'yawn' is unconnected.) *teguh TB togu 'sturdy, fast', Ja, Ml teguh 'firm and strong, tough', NgD tagoh (a borrowing) 'impenetrable, impervious to wounds'; *pagut TB pagut, pargut 'bill of a bird, mouth of a snake', Ml pagut 'peck (bird), bite (snake), nibble (fish)'. (Ja pagut 'take hold of' is not verified.) *tagam TB tagam 'be wary of something', Cebuano tagam 'have had a bad experience so that one does not wish to repeat it' (The similarity may be purely coincidental. Cebuano also has a form tagamtam 'experience' which is from a root *tamtam, and the form tagam may be derived from this. Dempwolff connects the TB form with *agam listed at the beginning of this section.) There is a single example of a *g in final position which cannot readily be explained away. I think that in this case we are dealing with a coincidence — that is, the Ml and Javanese forms are connected by borrowing (as the vocalism of the Ml form indicates), but these only coincidentally resemble a Tg form which has a similar meaning: *lunzag Ml lonjak, Jav lunjak 'spring up', Tg lundag 'jump, leap'.

4. *T

The evidence for *T is very slender. Dempwolff thought that the occurrence of Ja T where other languages have t (or a reflex of *t) indicated *T, but Ja was thought to be the only language that did not merge *t and *T. This fact alone makes *T suspicious. Further, we find that most of the forms in the correspondence *T are words subject to contamination (analogical reformation), a great number of them confined to only Ml and Ja, a few of them learned borrowings, a few not connected in meaning, and a few which could perfectly well be reconstructed with *t. This leaves four forms in this correspondence which could possibly be inherited. Phonemic conditioning or pure chance must account for these.

4.1. Many Ja forms which refer to sounds contain a phoneme T, whereas t practically fails to occur in onomatopoeic forms. The conclusion is inescapable that T has spread into forms of this type by an analogical process, and these forms can offer no evidence for a proto-phoneme. The following reconstructions by Dempwolff refer to forms of onomatopoeic meanings: *keTek, *keTik, *kenTung, *keTuk, *Tek, *TekTek, *Tik, *TikTik, *Tak, *TakTak, *Ting, *TingTing, *Tung, *Tuk, *TukTuk, *keTak. The form *keTer (Blust) which refers to shaking, trembling is also of the sort that is liable to analogic change (cf. the discussion in Section 3.1., above). The word for 'little' is also subject to analogical reformation of this sort: *iTik TB, Minangkabau ètèk, Ja Titik, iTik, Sm iti-iti 'little', Ho vitsi 'few' (not connected). In the first place the Ja form has an initial T, already a clue that analogical change has taken place. Further, there are numerous other forms with similar sounds in a similar meaning: Ml kecil, kecil 'little', etc. The Ja form cannot have been inherited. I believe that analogical reformation is also involved in the case of *peTik Ja peTik, Ml petik 'pluck, pick' (Fi peti 'pluck fruit' is not directly related.) The form *putik is well attested and the reflexes look inherited from the proto-language. (Dempwolff reconstructed *puTik, but there is no Ja cognate.) TB putik 'pluck', Fi vutia 'pluck out hair', PPN *futi 'pluck, pull out'. Ja peTik shows an irregularity of the vowel of the first syllable, and Ml petik probably developed from *putik, under the influence of Ja. In any case, the Ja does not reflect a proto-form directly.

4.2. A number of forms reconstructed with *T have the character of learned borrowings or are the names of flora or fauna: *baTik Tg batic 'painted cloth, tattoos (Noceda)', Ja baTik 'batik', NgD batik 'thick cloth with flowery patterns'; *keTem NgD, Tg katam 'plane' (with vocalism which proves borrowing), TB otom 'plane' (with an irregular initial sound), Ml ketam 'plane, grasp in claws', Ja keTem 'tongs to hold something in forging, grasp in claws of the like'; *kaTil Ja, katil, Ml, NgD katil 'bedstead' (This form is borrowed from Tamil kattil.) *keTu Ja keTu 'cap around which a turban is wound', Ml keto 'ascetic's headdress of Hindu period'; *banTing Old Ja banTéng 'water buffalo', Ml, NgD banting 'wild buffalo'. Other forms in addition to those above which are confined to Ja and Ml or to these and NgD are surely not related by inheritance from the proto-language: *buTak Ja buTak 'bald, bare', Ml botak 'bald, having an empty spot at crown of head'; *kanTung Ja kanTong, NgD kantong, Ml kantong 'pocket' (Fi kato, To kato and Sm 'ato 'basket' are not related.) *kenTel Ja kenTel, Ml kental 'thick, viscous'; *kenTang Ja kenTang, Ml, NgD kentang, NgD kantang 'potato' (Hardelandt explains that these were not known to the Ngajus at his time and says that the word he lists is the Ml word.) *kunTul Jav konTol 'penis, scrotum', Ml kuntul, kontrol, keruntul, kerental, kental (all referring to thick pendulous objects); *kuTung Ja kuTung, Ml kotong 'docked'; *paTi Ja paTi 'starch, essence', Ml pati 'cream, finest portion of anything' (To matsi 'scented, sweet smelling after having leaves soaked in it' is not connected.) *Tu(ng)kul Ja Tukul 'sprout', NgD tungkol 'banana bud' (Tg tukol 'even in number, overripe rice grains that are unharvested' is not connected.)

Some of the forms in the *T correspondence are patently unconnected in meaning: *aTik Ja aTik-aTikan 'something made up', TB atik 'perhaps', NgD etek, petek 'intent'; *baTuk Tg bAtok 'nape of neck', Old Ja baTuk 'forehead', Ja baTuk 'coconut shell', Ml batuk 'coconut shell' (said to be from Ja)', Sa qä'u 'head', Sm patu 'swelling, lump'; *beTak Tg bitak 'crack, cleft', Ja beTak 'boil rice', Fi beteka 'break, brittle things', Sa ho'a 'take aside, separate'; *bunTing Ja bunTing (Dempwolff's gloss 'youngest child' not confirmed), Ml bunting 'pregnant', TB butteng-butteng 'pregnant', Sa huti 'be born'; *buTek Ja buTeg 'water that is not clear, roiled up', Fi butoo 'darkness'; *hi(n)Tai Tg hintay 'wait', TB itte 'wait', Ja iTi 'take scrupulous care of', Ml intai 'spy on'. (The TB and Tg forms may be related, but they furnish no evidence for T.) *Tangkur Tg tangkol 'cuff, box', TB tahur 'hollow out a log', Ja Takur-Takur 'paw, dig with hands or feet', Ho tanggu 'stripped of wings and legs (locust), herbs stripped for cooking'; *Ta(r)ik Tg tálik 'intimate, close', Ja Tarik-Tarik 'set in neat rows'; *baTang Old Ja babaTang 'dead body', TB batang 'coffin'.

4.4. A few forms which have been reconstructed with *T can as well be reconstructed with *t: *anTuk Old Ja mangantuk (with t not T) 'sleepy, for the head to droop'; *DeTik Ml detik 'ticking sound, second', Tg lítik 'pulse beat' (no Ja form quoted); *piTik Tg pitik 'fillip' (and several other meanings referring to sudden motion), TB pitik 'thrown down, apart from others', NgD pitik 'quick, deft', Fi vidi 'jump, fly up', vidika 'fillip', PPN *fiti 'spring up, move suddenly, fillip' (Ja pèTèk 'shock of hair' has no connection.) *Tenguk Ja, Ml tèngok 'peer at', Fi digova 'examine, inspect' (There is no reason to reconstruct with *T.

4.5. This leaves us with four or possibly five forms which cannot be explained away. Two of them have cognates in languages outside of Indonesia and the Philippines. They are *puTul Tg pútol 'cut off', Ja puTul 'break off', To, Fu, Sm mutu 'cut off, ended'; *kiTa Tg kita 'see', Ja waskiTa 'with clear insight', NgD ite 'see', Ho hita 'see', Fi kida 'salute a person on arrival', kida-kida 'come on an enemy', PPN *kite 'see, appear, know'. In the case of puTul, it may well be that the Ja form happens to resemble the Philippine and Polynesian forms. Since there are endless numbers for forms for 'cut, break off' in these languages, many of which resemble each other in sound (giving rise to possibilities for analogical reformations), it can well be that the resemblance of these forms is purely fortuitous. In the case of *kiTa, it is not absolutely certain what the meaning of the Ja form -kiTa is in the compound waskiTa, although it probably does mean something like 'see'. However, even though reflexes of this form are found from Formosa to Oceania, they show irregularities of correspondence in many languages (e.g. the lack of initial k- in NgD, the final vowel of the Polynesian forms, Cebuano has kitaq with an unexplained glottal stop). This indicates that analogical changes which we are not in a position to understand have reshaped this form in many languages. Two other forms with *T are found only in Ja and one other language. Even though there is no Ml form attested for the correspondence, the meanings are such that they could well be related by borrowing, and in one case, doublets in Ja make this a virtual certainty: *piTung TB pitung 'blind', Ja piTo, piTong 'blind in one eye', Tg puntok 'cone, apex', Ja punTuk 'mound, hill, peak'.

NOTES

1. Dyen (1971) lists 25 of the correspondences which he finds may reflect phonemes which should be added to the PAN phonemic inventory.
2. The evidence for *Z, *N, and *C is very strong, and these were surely phonemes of the proto-language.
3. I adhere to the symbols for the correspondences which have become traditional, for ease of reference. However, an urgent task for Austronesian linguistics is to investigate the probable phonetic content of these phonemes on the basis of their reflexes.
4. To use a loanword as evidence for the existence of a proto-phoneme is to assert that the loanword was borrowed during the time of the proto-language, clearly a ridiculous thing to assert in the case of Indic loanwords, and not very likely in the case of Chinese forms either.
5. I am using Dempwolff's abbreviations for the languages: Ml: Malay, Ja: Javanese, Tg: Tagalog, TB: Toba Batak, NgD: Ngaju Dayak, Ho: Malagasy, Fi: Fijian, Sa: Sa'a, Sm: Samoan, To: Tongan, Fu: Futunan. I also use the abbreviations PPN for Proto-Polynesian, and PAN for Proto-Austronesian. In the case of this particular correspondence, Dempwolff thought NgD reflected *c with c, but Dyen (1956) showed that NgD forms with c were borrowings from Ml.
6. Dempwolff only thought that Ml and Ja had reflexes of *ns different from *nc (Ml and Ja ngs reflecting *ns and nc reflecting *nc). The number of forms with *ns is very small and the n is clearly a later accretion. First, the forms Dempwolff reconstructed with *ns on the basis of the Javanese reflex with ngs: *ha(n)sang, *ansu, *li(n)sa. In these three cases I believe the Javanese n to be a later accretion: Javanese has a propensity for inserting nasals at random after the vowel of the penult. These forms also occur with by-forms without the medial nasal, so that the Javanese forms cannot be taken to be evidence for a contrast between *ns and *nc. Second, forms with Malay ngs: *la(n)sat, *bu(n)su, *punsu. Here we are dealing with Malay words which have come into the other languages as borrowings. *La(n)sat is the name of a fruit. *Bu(n)su is attested by Tg bunsoq, Ml bungsu, and Ho busu. The final q in Tg proves that the form is a Tg borrowing from Ml, and the Ho form proves that the nasal is a recent development in Ml. *Pu(n)su is attested by Ml pungsu, or pusu, and Tg punso. The existence of the variant shows that the nasal most likely is secondary in Ml. The form in Tg is likely to be a borrowing, for the term is used in the phrase matandá sa punso, the name of a supernatural creature which corresponds to something in the Ml cosmogeny, and thus is likely to be a borrowing. Many of the terms for Philippine magic are borrowed from Ml magical terms. In short, there is no evidence for a contrast between *nc and *ns.

7. An example is kambana 'twin' (from Ml kembar 'twin'). The correspondence Ho a Ml e indicates a loanword. Another example is Ho ranto 'journey to distant places for commercial purposes'. This is a borrowing from Ml rantau, the basic meaning of which is 'the reach of a river or the far shore' but which by extension means 'go abroad to earn money'. We know this form is a borrowing in Ho because it has the extended meaning.
8. Gonda believes this to be borrowed from Sanskrit pūjā 'worship' in view of the Ja meaning.
9. In the case of tazem, the form tarem in Sambal (Philippines) must be a borrowing from the Cordilleran languages. The NgD form has an irregular reflex of the vowel of the final syllable. Thus, it is not unlikely that these forms are in fact spread by borrowing. In the case of *za(m)bay reflexes are attested only in TB and NgD. The meanings are close, and not of the sort to be borrowed, but it is the sort of thing that can be subject to contamination (analogical change) from forms with similar meanings, and I think the correspondence is coincidental.
10. Examples of this are Ja gendèra, Ml bendèra 'flag' (from Portuguese), Ml gelisah, belisah 'disquiet', Ja dandang 'steam', Ml ganggang 'heat', Ml gapas, Tg kapas 'cotton', etc.
11. The basic meaning of Ja gelar is 'spread out'. It also came to mean 'exposed to public, spread out to view' from which the meaning 'title, thing by which one is known' developed.

INDEX OF FORMS

Proto-Austronesian

- *agak 3.3.
 *agam 3.5.
 *ager 3.3.
 *anjaw 3.
 *ansu 1. note 6
 *anTuk 4.4.
 *anzang 2.3.
 *aTik 4.3.
 *azak 2.4.
 *azar 2.4.
 *azi 2.1.
 *bagus 3.3.
 *banTing 4.2.
 *banzar 2.4.
 *baTik 4.2.
 *baTuk 4.3.
 *bazi 2.1.
 *belaj 3.3.
 *beTak 4.3.
 *bu(n)su 1. note 6
 *bunTing 4.3.
 *buTak 4.2.
 *buTek 4.3.
 *cabat 1.4.
 *cabang 1.2.
 *cabut 1.5.
 *cacar 1.2.
 *cacing 1.2.
 *cagak 1.2.
 *cahing 1.3.
 *cakep 1.2.
 *cahuR 1.5.
 *calcal 1.4.
 *cambuk 1.2.
 *campur 1.5.
 *camuk 1.5.
 *candu 1.2.
 *ca(ng)kem 1.3.
 *cangkul 1.2.
 *ca(ng)kup 1.2.
 *canTik 1.2.
 *canTing 1.2.
 *canTung 1.2.
 *capah 1.2.
 *cecak 1.1.
 *ceguk 1.1., 3.1.
 *cekig 3.2.
 *celak 1.2.
 *cemcem 1.3.
 *cemeD 1.2.
 *cencen 1.4.
 *cenTung 1.2.
 *cepat 1.2.
 *ciap 1.1.
 *ci(m)pir 1.2.
 *cincin 1.2.
 *ci(n)Duk 1.5.
 *citcit 1.1.
 *ciTak 1.2.
 *ciyum 1.2.
 *cukup 1.2.
 *cu(ng)kil 1.2.
 *cuping 1.5.
 *degDeg 3.2.
 *deTik 4.4.
 *dugDug 3.4.
 *gaduk 3.2.
 *gaDay 3.2.
 *gaDgaD 3.2.
 *gaDung 3.3.
 *gagah 3.3.
 *gagak 3.1.
 *gahul 3.2.
 *galak 3.5.
 *galang 3.1.1.
 *galaq 3.5.
 *galing 3.1.
 *galgal 3.2.
 *galumat 3.3.
 *galumbang 3.3.
 *ga(m)pit 3.1.1.
 *gana 3.3.
 *gangsai 3.3.
 *ganggang 3.5.
 *ganggu 3.5.
 *gantai 3.3.
 *gantang 3.3.
 *ganti 3.3.
 *gantung 3.5.
 *ganzil 3.3.
 *gasang 3.2.
 *gasing 3.1.1.
 *gapay 3.5.
 *ga(r)ang 3.5.
 *gatel 3.1.1.
 *gawai 3.3.
 *gawa 3.3.
 *gayang 3.2.
 *gayung 3.3., 3.5.
 *geber 3.3.
 *gelang 3.3.
 *gelgel 3.2.
 *gembar 3.1.
 *gemgem 3.2., 3.5.
 *gemuk 3.3.
 *genDang 3.1.1.
 *genDit 3.1.1.
 *genep 3.3.
 *ge(n)tas 3.1.
 *genti 3.3.
 *genting 3.5.
 *genung 3.3.
 *geli 3.1.1.
 *gelap 3.5.
 *gelar 3.3.
 *gesa 3.2.
 *getaq 3.3.
 *gigi 3.2.
 *gigil 3.1.
 *gigit 3.3.
 *gila 3.3.
 *gilang 3.1., 3.2.
 *gilap 3.1.1.
 *giling 3.1.
 *gili(r) 3.3.
 *gilis 3.2.
 *gitik 3.1.1.
 *guDang 3.2.
 *gugut 3.2.
 *guham 3.3.
 *gukguk 3.1.
 *gulai 3.3.
 *guling 3.1.
 *guluk 3.3.
 *gulong 3.1.

Proto-Austronesian (cont)

- *gumi 3.1.
 *guncang 3.1.1.
 *guling 3.1.
 *gunDik 3.3.
 *gunDul 3.3.
 *gunting 3.3.
 *guRi 3.3.
 *gusar 3.3.
 *gusgus 3.2.
 *gusuk 3.1.1.
 *gutem 3.2.
 *guyang 3.1.
 *ha(n)sang 1. note 6
 *hi(n)Tai 4.3.
 *hizaw 2.3.
 *iTik 4.1.
 *kalang 3.1.1.
 *ka(m)pit 3.1.1.
 *ka(n)cang 1.2.
 *kanTung 4.2.
 *kaTil 4.2.
 *kazang 2.1.
 *kenTang 4.2.
 *kenTel 4.2.
 *kenTung 4.1.
 *keTak 4.1.
 *keTek 4.1.
 *keTem 4.2.
 *keTu 4.2.
 *keTuk 4.1.
 *keTer 4.1.
 *keTik 4.1.
 *ki(n)cu 1.2.
 *kiTa 4.5.
 *kucak 1.2.
 *kucup 1.3.
 *kumis 3.1.
 *kunTul 4.2.
 *kuTung 4.2.
 *la(n)sat 1. note 6
 (legung 3.1.
 *licak 1.1.
 *li(n)cad 1.3.
 *li(n)sa 1. note 6
 *lucut 1.2.
 *lunzag 3.5.
 *pagut 3.5.
 *panzi 2.1.
 *panzang 2.3.
 *paTi 4.2.
 *pecaq 1.3.
 *pecel 1.3.
 *pecut 1.2.
 *pegang 3.1.
 *peTik 4.1.
 *picek 1.2.
 *piTik 4.4.
 *piTung 4.5.
 *pucat 1.2.
 *pucuk 1.1.
 *punsu 1. note 6
 *puTik 4.1.
 *puTul 4.5.
 *puzi 2.1.
 *sagu 2.3.
 *sa(ng)gap 3.2.
 *saleng 1.3.
 *sigai 3.3.
 *segseg 3.4.
 *sugsug 3.2.
 *suci 1.2.
 *suligi 1.3.
 *suzi 2.1.
 *tagam 3.5.
 *ta(n)zak 2.1.
 *ta(n)zuk 2.4.
 *tanzung 2.1.
 *tazak 2.1.
 *tazi 2.1.
 *tazem 2.4.
 *tegend 3.3.
 *teges 3.2.
 *(t)eguk 3.1.
 *tibag 3.2.
 *ti(n)zak 2.3.
 *tinzaw 2.3.
 *tinzu 2.1.
 *Tak 4.1.
 *TakTak 4.1.
 *Tangkur 4.3.
 *Ta(r)ik 4.3.
 *Tek 4.1.
 *Tik 4.1.
 *TikTik 4.1.
 *Ting 4.1.
 *TingTing 4.1.
 *Tuk 4.1.
 *TukTuk 4.1.
 *Tung 4.1.
 *Tu(ng)kul 4.2.
 *ucap 1.2.
 *ulug 3.2.
 *uzi 2.1.
 *uzung 2.1.
 *zabi 2.1.
 *zagal 2.1.
 *zagzag 2.2.
 *zahat 2.3.
 *zahit 2.3.
 *zalang 2.1.
 *zalateng 2.3.
 *zalin 2.4.
 *zalzal 2.3.
 *zamah 2.4.
 *zambang 2.1.
 *za(m)bat 2.4.
 *za(m)bay 2.4.
 *zamzam 2.2.
 *zanggut 2.2.
 *zangka 2.2.
 *zanzi 2.4.
 *zaRum 2.3.
 *zatuh 2.1.
 *zava 2.2.
 *zawzaw 2.2.
 *zayu 2.4.
 *zehet 2.2.
 *zegzeg 2.2.
 *zelag 2.2.
 *zelai 2.2.
 *zelar 3.3.
 *zempuT 2.1.
 *zemuR 2.1.
 *zeng 2.2.
 *zengkal 2.1.
 *zinak 2.4.
 *zingat 2.2.
 *zingzing 2.2.
 *zinten 2.1.
 *zual 2.3.
 *zuling 2.3.
 *zuluk 2.4.
 *zulzul 2.1.
 *zu(m)put 2.1.
 *zungzung 2.1.
 *zu(ng)kir 2.2.
 *zu(ng)kuk 2.2.
 *zu(ng)kung 2.1.
 *ZalZal 2.3.
 *Zambat 2.4.
 *Zaqet 2.3.
 *Zaqit 2.3.
 *ZaRum 2.3.
 *Zayuq 2.4.
 *Zayqu 2.4.
 *Zual 2.3.
 *Zuling 2.3.

Proto-Polynesian

*fiti 4.4.
 *futi 4.1.
 *kite 4.5.
 *sapai 2.4.
 *saqi 2.3.
 *soko 2.2.

Cebuano

bitas 3.1.
 bugtas 3.1.
 dayqu 2.4.
 daqut 2.3.
 gábut 1.5.
 ganta 3.3.
 gitik 3.1.1.
 ibut 1.5.
 káhug 1.5.
 kasing 3.1.1.
 kitaq 4.5.
 láhug 1.5.
 pisaq 1.3.
 pisil 1.3.
 tagam 3.5.
 tagamtam 3.5.
 *tamtam 3.5.
 tindak 2.3.
 tupay 3.5.
 ulug-úlug 3.2.

Fijian

basa 2.4.
 beteka 4.3.
 butoo 4.3.
 caa 2.3.
 caita 2.3.
 cambaka 1.4.
 cavuta 1.5.
 cenga 2.2.
 coko 2.2.
 covuta 2.1.
 doka 3.2.
 digova 4.4.
 kaci 2.1.
 kato 4.2.
 kabita 3.1.1.
 kida 4.5.
 kida-kida 4.5.
 kiila 3.3.
 kilica 3.2.
 kucuva 1.3.
 kumi 3.1.
 kusa 3.3.
 namuta 1.5.

Fijian (cont)

peti 4.1.
 qesa 3.2.
 qili 3.1.
 qilica 3.1.
 qoma 3.5.
 qosa 3.2.
 saambaya 1.4.
 salato 2.3.
 sau 2.3.
 savirewa 2.1.
 somo 1.2.
 sosoo ucu 1.4.
 tacavu 1.5.
 tasi 2.4.
 tirova 2.3.
 voca 1.3.
 vosa 1.2.
 vili 4.4.
 vidika 4.4.
 vutia 4.1.
 ucu 2.1.

Futuna

kata 3.3.
 kikila 3.1.
 kuku 3.1.
 mutu 4.5.
 saka 2.1.

Javanese

agak 3.3.
 agar-agar 3.2.
 ager-ager 3.2.
 ajar 2.4.
 bagus 3.3.
 banjar 2.4.
 baTang 4.3.
 baTik 4.2.
 beTak 4.3.
 bunTing 4.3.
 buTak 4.2.
 buTeg 4.3.
 cabut 1.5.
 cacah 1.4.
 cacar 1.2.
 cacing 1.2.
 cagak 1.2.
 cakep 1.2.
 cakup 1.2.
 campur 1.5.
 camuk 1.5.
 cangkul 1.2.
 canTik 1.2.

Javanese (cont)

canTing 1.2.
 canTung 1.2.
 cawad 1.4.
 cawang 1.2.
 cawet 1.4.
 ceguk 1.1.
 cekik 3.2.
 celak 1.2.
 cemcem 1.3.
 cemer 1.2.
 cèng 1.3.
 cenTung 1.2.
 cepèr 1.2.
 cèpèr 1.2.
 cepet 1.2.
 ciDuk 1.5.
 cincin 1.2.
 ciTak 1.2.
 cium 1.2.
 cungkil 1.2.
 cukup 1.2.
 dol 2.3.
 dom 2.3.
 DegDeg 3.2.
 DoDop 3.4.
 gaDe 3.3.
 gaDung 3.3.
 gagah 3.3.
 gagak 3.1.
 gagal 3.2.
 gagar 3.2.
 galak 3.5.
 gana 3.3.
 gendèra 3. note 10
 ganep 3.3.
 ganggang 3.5.
 ganggu 3.5.
 ganjil 3.3.
 gantal 3.3.
 gantang 3.3.
 ganti 3.3.
 gape 3.5.
 gapit 3.1.1.
 garang 3.5.
 gawe 3.3.
 gas 3.2.
 gasingan 3.1.1.
 gatel 3.1.1.
 gayung 3.3.
 geber 3.3.
 gegel 3.2.
 gelang 3.3.
 gelap 3.5.

Javanese (cont)

- gelar 3.3.
 gemgem 3.5.
 gemuk 3.3.
 genep 3.3.
 genti 3.3.
 genting 3.5.
 geseng 3.2.
 getas 3.1.
 gigit 3.3.
 gila 3.3.
 gilang 3.1.
 gilap 3.1.1.
 giles 3.2.
 giling 3.1.
 gilir 3.3.
 gilis 3.2.
 geli 3.1.1.
 glenggem 3.2.
 gogos 3.2.
 golok 3.3.
 gom 3.3.
 gosok 3.1.1.
 goyang 3.1.
 gugut 3.2.
 guling 3.1.
 gunDik 3.3.
 gunting 3.3.
 gunung 3.3.
 gusar 3.3.
 gusi 3.3.
 ijo 2.3.
 iTi 4.3.
 iTik 4.1.
 jabut 1.5.
 jahat 2.3.
 jait 2.3.
 jak 2.4.
 jalang 2.1.
 jali 2.1.
 jamah 2.4.
 jamban 2.1.
 janggut 2.2.
 jangka 2.2.
 langkah 2.1.
 janji 2.4.
 jawatan 2.4.
 jeng 2.2.
 jinak 2.4.
 jinjing 2.2.
 jinten 2.2.
 jojo 2.2.
 jujul 2.1.
 jukuk 2.1.
 jukung 2.1.
 julukaké 2.4.
 jumput 2.1.
 jungkruk 2.2.
 jungkung 2.1.
 jupuk 2.1.
 kagol 3.2.
 kajang 2.1.
 kanTong 4.2.
 kaTil 4.2.
 kenDang 3.1.1.
 kenDit 3.1.1.
 kenTang 4.2.
 keTem 4.2.
 keTu 4.2.
 kilingan 3.1.
 -kiTa 4.5.
 kocak 1.2.
 konTol 4.2.
 kunting 3.3.
 kuping 1.5.
 kuTung 4.2.
 lincad 1.3.
 lunjak 3.5.
 nggukguk 3.1.
 njengat 2.2.
 panjang 2.3.
 panji 2.1.
 paTi 4.2.
 pecut 1.2.
 peTik 4.1.
 pagut 3.5.
 péTék 4.4.
 picek 1.2.
 piTo 4.5.
 piTong 4.5.
 pucet 1.2.
 pucuk 1.1.
 puji 2.1.
 punTuk 4.5.
 puTul 4.5.
 sagu 3.3.
 suci 1.2.
 suligi 3.3.
 susuk 3.2.
 taji 2.1.
 tajam 2.4.
 tanjak 2.1.
 tanjuk 2.4.
 tegeng 3.3.
 teges 3.2.
 tindak 2.3.
 tinjo 2.3.
 tuntas 3.1.
 Takur-Takur 4.3.
 Tarik-Tarik 4.3.
 Tenguk 4.4.
 TiTik 4.1.
 Tukul 4.2.
 ucap 1.2.
 uji 2.1.
 ujung 2.1.
 waskiTa 4.5.
 Old Javanese
 aji 2.1.
 babaTang 4.3.
 banTèng 4.2.
 cukil 1.2.
 gangsal 3.3.
 gemgem 3.5.
 guling 3.1.
 jagal 2.1.
 jagjag 2.2.
 jambe 2.1.
 kacang 1.2.
 kècu 1.2.
 mangantuk 4.4.
 suji 2.1.
 Malagasy
 andrandrana 2.3.
 busu 1. note 6
 ditsaka 1.1.
 fetsaka 1.3.
 fitsuka 1.2.
 gaga 3.1.
 halana 3.1.1.
 hantuna 3.5.
 hasina 3.1.1.
 hatina 3.1.1.
 hentana 3.1.
 hihi 3.2.
 hilana 3.2.
 henika 3.3.
 hita 4.5.
 hudina 3.1.
 huntsana 3.1.1.
 hurana 3.2.
 jelaka 2.2.
 juluka 2.4.
 juku 2.2.
 junjuna 2.1.
 kambana 1.2. note 7
 kava-kava 3.3.
 lutsaka 1.3.

Malagasy (cont)

kusuka 3.1.1.
 maitsu 2.3.
 mitsatsa 1.4.
 saku 3.3.
 sesitra 3.4.
 sufina 1.4.
 tanggu 4.3.
 tanjuna 2.1.
 tsaka 3.2.
 tsaku 1.2.
 tsamutsamuka 1.5.
 tsatsaka 1.1.
 tsetsetra 1.1.
 tsindruka 1.5.
 tsentsina 1.4.
 tsinjaka 2.3.
 tsinju 2.3.
 tsiutsiuna 1.2.
 tsuki 1.2.
 vitsi 4.1.
 zavi 2.1.
 zaitra 2.3.
 zehi 2.1.

Malay

agak 3.3.
 agar-agar 3.2.
 ajak 2.4.
 ajar 2.4.
 aji 2.1.
 anjang 2.3.
 bagus 3.3.
 baji 2.1.
 banjar 2.4.
 banting 4.2.
 batuk 4.3.
 baur 1.5.
 bècak 1.1.
 bècèk 1.1.
 belisah 3. note 10
 bendèra 3. note 10
 botak 4.2.
 bungsu 1. note 6
 bunting 4.3.
 cabang 1.2.
 cabut 1.5.
 cacar 1.2.
 cagak 1.2.
 cakap 1.2.
 cakup 1.2.
 cambuk 1.2.
 campur 1.5.
 candu 1.2.

Malay (cont)

cangkul 1.2.
 canting 1.2.
 cantung 1.2.
 capah 1.2.
 cawang 1.2.
 cawat 1.4.
 cecak 1.1.
 cegok 3.1.
 cekik 3.2.
 celak 1.2.
 cemar 1.2.
 cengkam 1.3.
 cepat 1.2.
 cèpèr 1.2.
 centung 1.2.
 cètak 1.2.
 ciap 1.1.
 cicit 1.1.
 ciduk 1.5.
 cincin 1.2.
 cium 1.2.
 cukup 1.2.
 cungkil 1.2.
 cuping 1.5.
 deguk 3.1.
 detik 4.4.
 gadai 3.3.
 gaduk 3.2.
 gadung 3.3.
 gagak 3.1.
 gagah 3.3.
 gagal 3.2.
 galah 3.5.
 galak 3.5.
 galang 3.1.1.
 ganggang 3. note 10
 ganggu 3.5.
 gangsang 3.2.
 ganjil 3.3.
 gantang 3.3.
 ganti 3.3.
 gantung 3.5.
 gapas 3. note 10
 gasal 3.3.
 gasang 3.2.
 gasing 3.1.1.
 garang 3.5.
 penggawa 3.3.
 gawai 3.3.
 gayang 3.2.
 gayung 3.3.
 gelar 3.3.
 gedang 3.2.

gelang 3.3.
 gelap 3.5.
 gelar 3.3.
 geli 3.1.1.
 gelisah 3. note 10
 gelitik 3.1.1.
 gelumat 3.3.
 gelumbang 3.3.
 gesa 3.2.
 gembar-gembor 3.1.
 gemuk 3.3.
 genap 3.3.
 gendang 3.1.1.
 gendit 3.1.1.
 genggam 3.5.
 gentas 3.1.
 genting 3.5.
 getah 3.3.
 getas 3.1.
 gigi 3.2.
 gigil 3.1.
 gigit 3.3.
 gila 3.3.
 gilang 3.1.
 gilap 3.1.1.
 giling 3.1.
 gilir 3.3.
 gokok 3.3.
 goncang 3.1.1.
 gondol 3.3.
 gosok 3.1.1.
 goyang 3.1.
 guam 3.3.
 gulai 3.3.
 guling 3.1.
 guncang 3.1.1.
 gundul 3.3.
 gunung 3.3.
 guri 3.3.
 gusar 3.3.
 gusi 3.3.
 hijau 2.3.
 intai 4.3.
 jabat 2.4.
 jabatan 2.4.
 jagal 2.1.
 jahat 2.3.
 jait 2.3.
 jalang 2.1.
 jalin 2.4.
 jamah 2.4.
 jamban 2.1.
 jambat 2.4.

Malay (cont)

- jambi 2.1.
 janji 2.4.
 jarum 2.3.
 jatuh 2.1.
 jauh 2.4.
 jangkah 2.1.
 janggut 2.2.
 jangka 2.2.
 jelatang 2.3.
 jemput 2.1.
 jemur 2.1.
 jènjèng 2.2.
 jengkal 2.1.
 jilai 2.1.
 jinak 2.4.
 jinjing 2.2.
 jintan 2.2.
 jojol 2.1.
 jolok 2.4.
 jongkok 2.2.
 jongkong 2.1.
 jual 2.3.
 jujul 2.1.
 juling 2.3.
 junjung 2.1.
 jukung 2.1.
 jungkang 2.1.
 kacang 1.2.
 kaju 2.1.
 kaji 2.1.
 kalang 3.1.1.
 kampit 3.1.1.
 kantung 4.2.
 katil 4.2.
 kecil 4.1.
 kecil 4.1.
 kecup 1.3.
 keliling 3.1.
 kendit 3.1.1.
 kental 4.2.
 kentang 4.2.
 kerental 4.2.
 keruntal 4.2.
 ketam 4.2.
 keto 4.2.
 kicu 1.2.
 kilap 3.1.1.
 kocak 1.2.
 koncang 3.1.1.
 kontal 4.2.
 kontol 4.2.
 kotong 4.2.
 kucup 1.3.
 kuntul 4.2.
 kuri 3.3.
 kumis 3.1.
 kuncang 3.1.1.
 melampai-lampai 3.5.
 lècah 1.1.
 lècak 1.1.
 lecap 1.1.
 lècèk 1.1.
 lècèt 1.1.
 legung 3.1.
 lèncèt 1.1.
 lincum 1.1.
 lonjak 3.5.
 lucut 1.2.
 olok 3.2.
 pagut 3.5.
 panjang 2.3.
 panji-panji 2.1.
 pati 4.2.
 pecah 1.3.
 pècak 1.2.
 pecal 1.3.
 pecut 1.2.
 petik 4.1.
 pucuk 1.1.
 pucat 1.2.
 puji 2.1.
 pungsu 1. note 6
 puncak 1.1.
 pusu 1. note 6
 ragok 3.1.
 rebut 1.5.
 retas 3.1.
 ruam 3.3.
 sagu 3.3.
 sampai 3.5.
 seligi 3.3.
 sigai 3.3.
 suci 1.2.
 bersuji 2.1.
 susuk 3.2.
 tajak 2.1.
 tajam 2.4.
 taji 2.1.
 tajuk 2.4.
 tanjak 2.1.
 tanjung 2.1.
 tebak 3.2.
 tegang 3.3.
 teguk 3.1.
 têngok 4.4.
 tèpèr 1.2.
 tinju 2.1.
 tinjau 2.3.
 ujung 2.1.
 ucap 1.2.
 uji 2.1.
 Minangkabau
 ètèk 4.1.
 Ngaju Dayak
 ajak 2.4.
 ajar 2.4.
 baji 2.1.
 banting 4.2.
 batik 4.2.
 cakap 1.2.
 bacakop 1.2.
 campur 1.5.
 capat 1.2.
 caur 1.5.
 etek 4.3.
 gambar 3.3.
 gadai 3.3.
 gadung 3.3.
 gagah 3.3.
 gagas 3.2.
 gagak 3.1.
 gagarang 3.5.
 gagoyang 3.1.
 gaguling 3.1.
 galing 3.1.
 galak 3.5.
 galang 3.1.1.
 galombang 3.3.
 galomat 3.3.
 gana 3.3.
 gandang 3.1.1.
 ganggo 3.5.
 ganjil 3.3.
 gantang 3.3.
 ganti 3.3.
 gantong 3.5.
 gargar 3.2.
 gatel 3.1.1.
 gawayan 3.3.
 gayang 3.2.
 gayong 3.3.
 gelang 3.3.
 kagenep 3.3.
 sagemgem 3.5.
 genep 3.3.
 getas 3.1.
 getem 3.2.

Ngaju Dayak (cont)

gila 3.3.
 gilap 3.1.1.
 giling 3.1.
 ginteng 3.5.
 girgir 3.3.
 gita 3.3.
 goam 3.3.
 goayan 3.3.
 gogot 3.2.
 gugut 3.2.
 gulai 3.3.
 gumi 3.1.
 gundik 3.3.
 gundul 3.3.
 gunong 3.3.
 gunting 3.3.
 hegang 3.1.
 hijau 2.3.
 ite 4.5.
 jajak 2.2.
 jalin 2.4.
 jao-jaoi 2.2.
 janggut 2.2.
 jangka 2.2.
 janji 2.4.
 jamban 2.1.
 jato 2.1.
 jawe 2.2.
 jawei 2.4.
 jilai 2.1.
 jinak 2.4.
 jintan 2.2.
 jual 2.3.
 julok 2.4.
 jukung 2.1.
 jumput 2.1.
 kacang 1.2.
 kapucok 1.1.
 kacacar 1.2.
 kagi 2.1.
 kajang 2.1.
 kalien 3.1.1.
 kantong 4.2.
 kantong 4.2.
 kasang 1.2.
 katam 4.2.
 katil 4.2.
 kentang 4.2.
 kitik 3.1.
 kuling 3.1.
 kunting 3.3.
 kusok 3.1.1.
 lusut 1.2.
 musat 1.2.
 nyanyantik 1.2.
 pacah 1.3.
 pagang 3.1.
 panggawa 3.3.
 panjang 2.3.
 pegang 3.1.
 petek 4.3.
 pitik 4.4.
 puji 2.1.
 sagak 1.2.
 sago 3.3.
 saing 1.3.
 sakap 1.2.
 basakop 1.2.
 mansakop 1.2.
 samar 1.2.
 sampur 1.5.
 sando 1.2.
 sangkul 1.2.
 saur 1.5.
 kasisit 1.1.
 sium 1.2.
 sukup 1.2.
 samput 2.1.
 sungkil 1.2.
 tagoh 3.5.
 tajak 2.1.
 taji 2.1.
 tajim 2.4.
 tajok 2.4.
 tampusok 1.1.
 tasak 1.1.
 tanjong 2.1.
 talsal 1.4.
 tisin 1.2.
 tungkol 4.2.
 Sa'a
 'äli 3.1.
 ata-ata 2.1.
 ho'a 4.3.
 huti 4.3.
 'iro 2.3.
 manata'ini 2.4.
 potu-potu 1.3.
 qä'u 4.3.
 suhi 1.5.
 sulitelu 2.3.
 toha 1.2.

Samoan

'a'a 3.1.
 'a'asa 3.2.
 'apili 3.1.1.
 'ata 3.3.
 'ato 4.2.
 'i'ila 3.1.
 iti-iti 4.1.
 motu 4.5.
 'oli-'oli 3.1.1.
 'o'omi 3.2., 3.5.
 patu 4.3.
 sa'a 2.1.
 salato 2.3.
 samu 1.5.
 sasa 2.3.
 sau 'ai tangata 2.4.
 somo 1.2.
 sulu 2.4.
 tapa-tapa 2.1.
 'u'u 3.1.

Tagalog

agam-ágam 3.5.
 áral 2.4.
 baling 1.5.
 bátik 4.2.
 bunsoq 1. note 6
 dait 2.3.
 daldal 2.3.
 dama 2.4.
 damdam 2.2.
 damput 2.1.
 dangkal 2.1.
 dayo 2.4.
 dáwa 2.2.
 dilag 2.2.
 dukit 2.2.
 duldul 2.1.
 duling 2.3.
 dulok 2.4.
 gadgad 3.2.
 gaga 3.3.
 gahol 3.2.
 galak 3.5.
 galang 3.3.
 galimgim 3.2.
 ganap 3.3.
 garil 3.3.
 gansal 3.3.
 ganti 3.3.
 gawai 3.3.
 gawaq 3.3.
 gasgas 3.2.

Tagalog (cont)

gataq 3.3.
 gigil 3.1.
 giling 3.1.
 gilgil 3.2.
 gimbal 3.1.
 gingting 3.5.
 gisa 3.2.
 gisa 3.2.
 guham 3.3.
 gukguk 3.1.
 gulai 3.3.
 gulang 3.2.
 gulok 3.3.
 gumi 3.1.
 gunting 3.3.
 gusgus 3.2.
 gutom 3.2.
 hilaw 2.3.
 hintay 4.3.
 kápit 3.1.1.
 kápas 3. note 10
 karang 2.1.
 karáyom 2.3.
 katam 4.2.
 kati 3.1.1.
 kilap 3.1.1.
 kíta 4.5.
 kunting 3.3.
 lagok 3.1.
 lagong 3.1.
 láyoq 2.4.
 liglig 3.2.
 linsad 1.3.
 lítik 4.4.
 luglug 3.4.
 lundag 3.5.
 lusot 1.2.
 pandi-pandi 2.1.
 pangadyiq 2.1.
 pisaq 1.3.
 pisak 1.2.
 pisil 1.3.
 pitik 4.4.
 punso 1. note 6
 puntok 4.5.
 puri 2.1.
 pútol 4.5.
 sabang 1.2.
 sabat 1.4.
 sagap 3.2.
 sagitsit 1.1.
 sago 3.3.
 sáhing 1.3.

sahog 1.5.
 sakam 1.3.
 sakop 1.2.
 salasaq 1.4.
 salsal 1.4.
 salugsug 3.2.
 samuk 1.5.
 sapol 1.5.
 sigok 1.1.
 sigsig 3.4.
 sigay 3.3.
 sikig 3.2.
 sílok 1.5.
 simsim 1.3.
 singsing 1.2.
 sinsin 1.4.
 siyap 1.1.
 sulígiq 3.3.
 suri 2.1.
 masúsiq 1.2.
 tagok 3.1.
 tálik 4.2.
 talim 2.4.
 tangkol 4.3.
 tárak 2.1.
 táriq 2.1.
 tibag 3.2.
 togok 3.1.
 túkol 4.2.
 úlok 3.2.
 úriq 2.1.
 úsap 1.2.
 Toba Batak
 agak 3.3.
 agam 3.5.
 agar-agar 3.3.
 ajar 2.4.
 aji 2.1.
 baji 2.1.
 ban jir 2.4.
 batang 4.3.
 butteng-butteng 4.3.
 gade 3.2.
 gaduk 3.2.
 gadong 3.3.
 gaga barani 3.3.
 gagak 3.1.
 gala 3.5.
 galak 3.5.
 galgal 3.2.
 galumbang 3.3.
 gamuk 3.3.

Toba Batak (cont)

gana 3.3.
 gang 3.5.
 ganggang 3.3.
 ganggu 3.5.
 ganjil 3.3.
 ganup 3.3.
 margao 3.3.
 gape 3.5.
 gappit 3.1.1.
 garang 3.5.
 gargar 3.2.
 gas 3.2.
 gasang 3.2.
 gasing 3.1.1.
 gatal 3.1.1., 3.3.
 gattang 3.3.
 gattung 3.5.
 gaung 3.5.
 geang-geang 3.2.
 gigi 3.2.
 gila 3.3.
 gilang 3.2.
 giling 3.1.
 gitik 3.1.1.
 goar 3.3.
 godang 3.2.
 golang 3.3.
 golap 3.5.
 goluk 3.3.
 gomgom 3.5.
 gondang 3.1.1.
 gondit 3.1.1.
 gonop 3.3.
 gosa 3.2.
 gosok 3.1.1.
 gota 3.3.
 gotti 3.3.
 gotting 3.5.
 guam 3.3.
 gukguk 3.1.
 gule 3.3.
 guling 3.1.
 gumis 3.1.
 guri-guri 3.3.
 gusar 3.3.
 gusgus 3.2.
 gutom 3.2.
 gutsang 3.1.1.
 hacang 1.2.
 hajang 2.1.
 halang 3.1.1.
 haling 3.1.
 hapit 3.1.1.

Toba Batak (cont)

hitsu 1.2.
 hilap 3.1.1.
 itte 4.3.
 jabatan 2.4.
 jabi-jabi 2.1.
 jagal 2.1.
 jait 2.3.
 jakjak 2.2.
 jakka 2.2.
 jalang 2.1..
 jalin 2.4.
 jaljal 2.3.
 jama 2.4.
 jamban 2.1.
 jambe 2.4.
 jamjam 2.2.
 janggut 2.2.
 janji 2.4.
 jarum 2.3.
 jat 2.3.
 jau 2.4.
 jehet 2.2.
 jinak 2.4.
 jingjing 2.2.
 jintan 2.1.
 jokkal 2.1.
 jomur 2.1.
 jongjong 2.2.
 joppot 2.1.
 jual 2.3.
 jukkit 2.2.
 juljul 2.1.
 julluk 2.4.
 lisat 1.3.
 lusak 1.3.
 lusut 1.2.
 otom 4.2.
 pagut 3.5.
 panji-panji 2.1.
 pargut 3.5.
 pitik 4.4.
 pitung 4.5.
 posa 1.3.
 puji 2.1.
 pulo tanjung 2.1.
 pusuk 1.1.
 putik 4.1.
 putsu 1.1.
 rogang 3.1.
 sabang 1.2.
 sapu 3.3.
 sandu 1.2.
 saor 1.5.

sapa 1.2.
 sappur 1.5.
 sarbut 1.5.
 sasa 1.4.
 suping 1.5.
 sippir 1.2.
 masoksok 3.4.
 somor-somor 1.2.
 parsottingan 1.2.
 sosak 1.1.
 sotson 1.4.
 suligi 3.3.
 sutsi 1.2.
 tagam 3.5.
 tahur 4.3.
 tajak 2.1.
 taji 2.1.
 tajom 2.4.
 tindo 2.3.
 tugong 3.3.
 togos 3.2.
 manguji 2.1.
 ujung 2.1.
 mangusaphon 1.2.

Tongan

haha 2.3.
 haka 2.1.
 hamu 1.5.
 hau 2.3.
 kaka 3.1.
 kakaha 3.2.
 kata 3.3.
 kato 4.2.
 kikila 3.1.
 koko 3.2.
 kokomi 3.5.
 lulu-lulu 3.4.
 matsi 4.2.
 mutu 4.5.

Ulawa

drämu 1.5.

BIBLIOGRAPHY

BLUST, Robert A.

- 1970 Proto-Austronesian addenda. *Oceanic Linguistics* 9/2:104-162.
- 1972 Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages'. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 4/8:1-17.
- 1973 Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages' - 2. *WPLUH* 5/3:33-61.

CLERCQ, F.S.A. de

- 1927 *Nieuw plantkundig woordenboek voor Nederlandsch Indië...* Amsterdam: J.H. de Bussy. (1st edn 1909.)

DEMPWOLFF, Otto

- 1934- *Vergleichende Lautlehre des austronesischen Wortschatzes.* Berlin:
1938 Dietrich Reimer.

DYEN, Isidore

- 1947 The Tagalog reflexes of Malayo-Polynesian D. *Language* 23:227-238.
- 1951 Proto-Malayo-Polynesian *Z. *Language* 27:534-540.
- 1953a *The Proto-Malayo-Polynesian laryngeals.* William Dwight Whitney Linguistic Series 9. Baltimore: Linguistic Society of America.
- 1953b Dempwolff's *R. *Language* 29:359-366.
- 1956 The Ngaju-Dayak "old speech stratum". *Language* 32:83-87.
- 1971 The Austronesian languages and Proto-Austronesian. In: Thomas A. Sebeok, ed. *Current trends in linguistics, vol.8: Linguistics in Oceania*, 5-54. The Hague: Mouton.

GONDA, J.

- 1952 *Sanskrit in Indonesia.* Sarasvati Vihara Series 28. Nagpur, India: International Academy of Indian Culture.

GRACE, George W.

- 1969 A Proto-Oceanic finder list. *WPLUH* 1/2:39-84.

LI, Paul Jen-kuei

- 1972 On comparative Tsou. *Bulletin of History and Philology, Academia Sinica* 44:311-338.

WALSH, D.S. and Bruce BIGGS

- 1966 *Proto-Polynesian word list I.* Te Reo Monographs. Auckland: Linguistic Society of New Zealand.

WOLFF, John U.

- 1974 Proto-Austronesian *r and *d. *Oceanic Linguistics* 13:77-121.

DICTIONARIES CONSULTED

Cebuano

WOLFF, John U.

- 1972 *A dictionary of Cebuano Visayan*. Data Paper no.87. Southeast Asia Program, Cornell University, Linguistic Series 6. Ithaca, New York: Cornell University.

Fijian

CAPELL, A.

- 1957 *A new Fijian dictionary*. Glasgow: Wilson Guthrie.

Futunan

GRÉZEL, S.M.

- 1878 *Dictionnaire futunien-français*. Paris: Maisonneuve.

Javanese

GERICKE, J.F.C. and Taco ROORDA

- 1901 *Javaansch-Nederlandsch handwoordenboek*. Amsterdam: J. Müller.

HORNE, Eleanor C.

- 1972 *A Javanese-English dictionary*. New Haven: Yale University Press.

PIGEAUD, Theodor

- 1938 *Javaans-Nederlands handwoordenboek*. Batavia: J.B. Wolters.

Old Javanese

MARDIWARSITO, L.

- 1978 *Kamus Jawa Kuna (Kawi) - Indonesia*. Ende, Flores: Nusa Indah.

Malagasy

ABINAL, P. and P. MALZAC, S.J.

- 1963 *Dictionnaire malgache-français*. Paris: Editions Maritimes et d'Outre-mer.

RICHARDSON, James

- 1885 *A new Malagasy-English dictionary*. Antananarivo: The London Missionary Society. Republished 1967 by Gregg Press, Farnborough, Hants., England.

Malay

ECHOLS, J. and H. SHADILY

1963 *An Indonesian-English dictionary*. 2nd edn. Ithaca, New York: Cornell University Press.

ISKANDAR, Teuku

1970 *Kamus Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

WILKINSON, R.J.

1932 *A Malay-English dictionary (romanized)*. Mytilene, Greece: Salavopoulos and Kinderlis.

Ngaju-Dayak

HARDELAND, August

1859 *Dajacksch-deutsches Wörterbuch*. Amsterdam: Frederik Muller.

Sa'a

IVENS, Walter G.

1929 *A dictionary of the language of Sa'a (Mala) and Ulawa, South-East Solomon Islands*. London: Oxford University Press.

Samoan

Milner, G.B.

1966 *Samoan dictionary*. London: Oxford University Press.

Toba Batak

WARNECK, Johannes G.

1977 *Toba Batak-Deutsches Wörterbuch*, ed. by R. Roolvink. The Hague: Martinus Nijhoff.

Tongan

CHURCHWARD, C.M.

1959 *Tongan dictionary*. London: Oxford University Press.

Tagalog

LAKTAW, Pedro S.

1914 *Diccionario Tagalog-Hispanico*. Manila: Santos y Bernal.

NOCEDA, P. Juan de

1860 *Vocabulario de la Lengua Tagala*. Manila.

PANGANIBAN, José Villa

1972 *Diksyunaryo tesauro Pilipino-Ingles*. Quezon: Manlapaz.

THE PRESENT STATUS OF SOME AUSTRONESIAN SUBGROUPING HYPOTHESES

Isidore Dyen

It has been suggested (Blust 1980:208) that the non-Formosan languages can be classified under the name Malayo-Polynesian into three co-ordinate groupings as Western Malayo-Polynesian (WMP), Central Malayo-Polynesian (CMP), and Eastern Malayo-Polynesian (EMP). The languages of the Philippines and of western Indonesia as far east as Sumbawan inclusive are assigned to WMP, and the Oceanic languages together with Bulic (the languages of South Halmahera and nearby islands closely related to Buli) and Biakic (the languages of Cenderawasih Bay closely related to Biak) are assigned to EMP. The languages in between these two linguistic groupings, i.e. those from Bima on Sumbawa, eastward at least to Kuiwai on the western coast of Irian are assigned to Central Malayo-Polynesian.

I have already dealt with some of the faults with the Eastern Malayo-Polynesian hypothesis (Dyen 1978) proposed by Blust (1978). In this paper I will treat the evidence bearing on the proposed seam between WMP and CMP said to fall between Sumbawan and Biman and compare it with the suggested seams in eastern Indonesia.

THE BALIC LANGUAGES

Balinese, Sasak, and Sumbawan show their highest lexicostatistical percentages with each other: Balinese-Sasak 39.4%, Sasak-Sumbawan 49.7%, and Balinese-Sumbawan 36.1%. Javanese, which is known to have had a sociopolitically dominant position relative to Balinese in Bali, shows 35.6% with Balinese, but 31.0% with Sasak and 30.5% with Sumbawan. Taking the relatively higher Javanese score with Balinese to be due to intimate borrowings from Javanese, we regard the Javanese-Balinese score as inflated. At the same time we infer that the scores of Balinese with Sasak and Sumbawan were probably somewhat depressed because some cognates in Balinese had been replaced by intimate borrowings from Javanese.

The finding of a Balic subgroup on a lexicostatistical basis would explain the agreement of Balinese and Sasak in being the only languages exhibiting the suffixes *-in* (roughly equivalent to Malay *-i*) and *-an* (roughly equivalent to Malay *-kan*). These suffixes are perhaps absent in Sumbawan; I have only noted the verb *bean* 'give' which is probably to be connected with Javanese *weh-* 'give', though both are probably in some way to be associated with **beRey*. However the

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 31-35. *Pacific Linguistics*, C-75, 1982.

© Isidore Dyen

31

exceedingly numerous vocabulary agreements between Sasak and Sumbawan that are difficult to explain as either cultural or intimate borrowings support the membership of Sumbawan in the Balic subgroup.

THE IMMEDIATE EXTERNAL RELATIONSHIPS OF BALIC

The Balic subgroup, including Sumbawan, was recognised by Esser (1938) as the 'Bali-Sasak Group'. Its percentages with Javanese might appear to suggest that the immediate relationship of Balic is to be sought to its west. The percentages of Balic with Bima, the eastern neighbour of Sumbawan, also on Sumbawa, are as follows:

Bal-Bim	25.4%
Sas-Bim	32.5%
Sbw-Bim	32.4%

If we regard the Balinese percentages as depressed by intimate borrowings, it seems to turn out that the Balic languages are about as distant from Bima as they are from Javanese, for the Balic-Bima relationship could now reasonably be estimated to be at about 30%.

If one considers the nearest relationships of Bima other than the above, we find the following percentages of Bima with its neighbours to the south and west, Waijewa (Kabhuhaka, Sumba), Waingapu (Kamera, Sumba), Sawu, and Manggarai (West Flores) as follows:

Bim-Wjw	25.4%
Bim-Wng	28.6%
Bim-Sawu	26.9%
Bim-Mng	23.8%

Curiously all of these percentages are slightly below the percentages that Bima shows with the Balic languages, if one excepts Balinese itself.

The speech types of Sumba seem to form a single linguistic grouping. Waijewa and Waingapu, probably among the most divergent pairs of Sumban speech types, show 44.1%. Sawu is favoured by the Sumban speech types but the precise relationship is indicated somewhat unclearly in the following percentages:

Sawu-Wjw	39.5%
Sawu-Wng	37.3%

If we considered only the Wjw-Wng percentage with each other, we should be unable to distinguish Sawu from another (perhaps somewhat aberrant) Sumban speech type. But the chances are that the chain of Sumban speech types has no link as low as the Wjw-Wng percentage. If that is so, then Sawu is lexicostatistically distinct from, but quite closely related to Sumban, probably more closely than are the speech types of Flores: Manggarai, Lio, Ngadha, Sika, and Lamaholot (= Solorese). The following table shows the lexicostatistical percentages shown by the relevant pairs of languages along with those with Balic and Bima:

Bal												
39.4	Sas											
36.1	49.7	Sbw										
25.4	32.5	30.7	Bim									
24.6	30.8	30.2	25.4	Wjw								
25.0	30.6	29.9	28.6	44.1	Wng							
24.5	28.7	28.8	26.9	39.6	37.3	Saw						
25.4	28.2	28.2	23.4	33.3	26.9	26.0	Wng					
22.0	24.4	25.8	25.9	31.6	27.7	27.7	41.1	Lio				
23.9	25.6	26.5	25.6	34.1	31.1	27.4	40.6	56.6	Ngdh			
26.0	30.9	32.0	27.4	36.5	33.2	30.4	36.9	42.7	39.0	Sik		
22.2	31.6	26.8	26.8	31.4	30.7	27.3	29.3	28.7	29.4	38.6	Lmh	

Table 1: Lexicostatistical percentages between languages on Bali, Lombok, Sumbawa, Sumba, Sawu and Flores

Pau						
25.0	Buli					
23.8	24.2	Biak				
16.0	15.5	18.5	War			
19.1	16.0	17.5	14.8	Sob		
23.5	17.8	20.7	16.1	58.5	Wak	

Table 2: Lexicostatistical percentages of six languages of Eastern Indonesia

BULIC, BIAKIC AND OTHER LANGUAGES OF EASTERN INDONESIA

A similar table has been drawn up for languages at the eastern end of Indonesia. The languages of Seram are represented by the thus far best described language there, Paulohi. Buli and Biak are included, the data for the latter being taken from the Van Hasselt dictionary of Numfor and my own recording of a dialect of Biak. The Waropen data is based on Held's work whereas the Sobey work is taken from Robert and Joyce Sterner's Sobey dictionary manuscript and my own recordings. The Wakde material is from my own work.

What seems to be most striking is that Buli and Biak exhibit the same scores with Paulohi as they do with each other. Thus there is no lexicostatistical evidence for a Bulic-Biakic that does not also include Paulohi and thus other languages of Seram closely related to Paulohi. Furthermore Biak shows about the same reaction to Sobey (17.5%) as the geographically much more distant Buli (16.0%). The same appears to be true if Wakde replaces Sobey as the focus.

It is interesting to observe that Waropen does not react like a submember of the same group as Biak. It may be because the study was too conservatively handled. The procedure that was followed was essentially the same as that of Dyen 1965 with one exception. Wherever one of a pair of comparable words was known to be a borrowing, the pair was regarded as non-comparable rather than as negative (as was done in Dyen 1965). However, the difference in treatment is believed to have negligible effects.

Furthermore the overall study of the phonemics of the different languages involved is now believed to be high enough to produce reliable results. What is perhaps weakest is the study of the correspondences. Needless to say, it is my belief that the relatively obvious instances have been properly counted.

The low level of the percentages in the eastern set is in keeping with the indications of Dyen 1965. This level suggests that the seam or seams involving Bima are between linguistic subgroupings more closely related to each other than the seams involving Paulohi, Buli, and Biakic with other eastern languages in their neighbourhood.

The lexicostatistical evidence favours the association of the languages in the neighbourhood of Bima with the so-called WMP at the clearly higher level of percentages than that associating Paulohi with Buli and Biak. It follows that if both Paulohi and Bima are in CMP, then CMP is more closely related to WMP than to EMP. Furthermore the evidence continues to point to an equivalently close relationship between Paulohi, Buli, and Biak, thus militating against the EMP hypothesis.

We seem to be able to conclude that the western seams involving Bima are probably of a lower order than the one between Sobey and Biak. What lies behind the relatively high percentage between Wakde and Paulohi (23.5%) remains to be investigated.

It is worth observing at this point that only lexicostatistical percentages permit the objective comparison of heterolinal nodes. Traditionally the comparison of heterolinal nodes has been made intuitively, but generally on the basis of large and practically incontrovertible masses of evidence pointing in the same direction. The family tree of the Austronesian languages seems to differ very strongly from that of the Indo-European languages not only by the multiplicity of its linguistic groupings, but also by the multiplicity of the linguistic groupings which do not lend themselves easily to subgrouping whether by traditional procedures or by lexicostatistics. The indications are thus strengthened that is is a family of great age whose classification will not yield quickly to intuitive pronouncements and will provide a challenge for many years to come.

BIBLIOGRAPHY

BLUST, Robert A.

- 1978 Eastern Malayo-Polynesian: a subgrouping argument. In: Wurm and Carrington, eds 1978:181-234.
- 1980 Early Austronesian social organization: the evidence of language. *Current Anthropology* 21:205-247.

DYEN, Isidore

- 1965 A lexicostatistical classification of the Austronesian languages. *International Journal of American Linguistics* Memoir 19. Baltimore.
- 1978 The position of the languages of eastern Indonesia. In: Wurm and Carrington, eds 1978:235-254.

ESSER, S.J.

- 1938 Talen. In *Atlas van Tropisch Nederland*, 9-96. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

HASSELT, J.L. van and F.J.F. van HASSELT

- 1947 *Noemfoorsch woordenboek*. Amsterdam: J.H. de Bussy.

HELD, A.J.

- 1942 *Grammatica van het Waropensch (Nederlandsch Noord Nieuw-Guinea)*. *Verhandelingen van het Bataviaasch Genootschap...* (VBG) 77/1.

MAAN, G.

- 1940 *Boelisch-Nederlandsche woordenlijst met Nederlandsch-Boelisch register*. VBG 74/3. Bandung: A.C. Nix.

STERNER, Joyce K. and Robert H. STERNER

- n.d. *Sobei dictionary*. MS, data collected 1973-1975, Irian Jaya.

STRESEMANN, Erwin

- 1918 *Die Paulohisprache: ein Beitrag zur Kenntnis der amboinischen Sprachengruppe*. The Hague: Nijhoff.

WURM, S.A. and Lois CARRINGTON, eds

- 1978 *Second International Conference on Austronesian Linguistics: proceedings*. PL, C-61.

A SYNTACTIC MODEL FOR THE COMPARATIVE STUDY OF AUSTRONESIAN LANGUAGES

Teodoro A. Llamzon

The purpose of this study is to find out whether a syntactic model could be devised which would provide a framework for the description of the syntactic structures of Austronesian languages. If so, such a model could also provide a basis for the comparative study of this family of languages, be it classification or reconstruction, since description of languages must necessarily precede most types of comparative work. (Anttila 1972:3).

This paper is divided into three parts. Part I will discuss the grammatical model used to describe Tagalog, the language on which Pilipino, the National Language of the Philippines, is based. Part II will show how the model is applicable to a number of Austronesian languages. Finally, Part III will point out the advantage of doing comparative work on languages whose structures have been described according to the same model.

SYNTACTIC STRUCTURE OF TAGALOG

In his book *Language*, Bloomfield made the following comments:

The era of exploration brought a superficial knowledge of many languages. Travelers brought back vocabularies, and missionaries translated religious books into the tongues of newly-discovered countries. Some even compiled grammars and dictionaries of exotic languages. Spanish priests began this work as early as in the sixteenth century; to them we owe a number of treatises on American and Philippine languages. These works can be used only with caution, for the authors, untrained in the recognition of foreign speech-sounds, could make no accurate record, and, knowing only the terminology of Latin grammar, distorted their exposition by fitting it into this frame. Down to our own time, persons without linguistic training have produced work of this sort; aside from the waste of labor, much information has in this way been lost. (1933:7)

In other words, Bloomfield's working principle in linguistic description

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 37-46. *Pacific Linguistics*, C-75, 1982.
© Teodoro A. Llamzon

is to account for the structures of a given language *on its own terms*. One should not try to 'distort' the description of a language by making its structures fit a certain grammatical model, but instead devise a model that would account for the structures of the language.

Following this lead, I have tried to see if first of all a syntactic model could be devised which would account for the structures of my mother tongue, Tagalog. If found adequate for Tagalog, it might then be possible to use the same framework for the description of other Philippine languages. The result of my work was published in Llamzon 1976, which contained a description of the phonology, morphology and syntax of the language. Here I would like to summarise briefly the structures on the sentence level, limiting the discussion to the basic sentence types of Tagalog.

There are at least three basic sentence types in Tagalog, namely, the monadic, diadic and triadic solidarities. The fourth type, quadradic solidarity, with four obligatory constituents, is not well attested. Examples of monadic solidarity (with one obligatory constituent) are: lumílingól '*there is an earthquake*' and may táo '*there is someone*' (the opposite of walang táo '*there is no one*').

Diadic solidarity (with two obligatory constituents) is illustrated by lumalákad akó '*I am walking*', si pédro ang pulís '*Pedro is the policeman*', pulá ang bulaklák '*the flower is red*', and maágang kumáin si pédro '*Pedro ate early*'.

Triadic solidarity (with three obligatory constituents) is exemplified by nagpatáy nang manók ang iná '*the mother killed a chicken*', naging sundálo si pédro '*Pedro became a soldier*', nagpumílit ang batang sumáma '*the child insisted on coming along*', kasinglaki ni huwán si pédro '*Pedro is as big as Juan*', maaáring malí si hosé '*Jose may be wrong*', and gusto kong kumáin nang manggá '*I want to eat a mango*'.

Quadradic solidarity (with four obligatory constituents) is not well attested, and the following sentence is considered as triadic by some Tagalog speakers but as quadradic by others: nagbúhos ang táo nang túbig sa bangáq '*the man poured water into the jar*'.

All the four types of solidarities may be expanded by the addition of adverbial phrases, noun phrases functioning as indirect object or as a noun phrase in the possessive case, and by the insertion of prepositives and/or enclitics. It is clear that this classification of sentence types is based on form rather than on function, and that the assignment of sentence structures to one of the four basic sentence types is based on the obligatory rather than the optional constituents of the sentence.

The framework presented here differs radically from that used by Bloomfield and other linguists, including more recent ones. In his book *Tagalog texts with grammatical analysis* (1917), Bloomfield employed a grammatical framework that was essentially diadic, which required a subject and predicate for every sentence, except the exclamatory and impersonal-anaphoric types. While such a framework was suitable for Indo-European languages, it certainly was not adequate for Tagalog. For example, in the light of the subject-predicate framework he employed, Bloomfield analysed expressions like umúulán '*it is raining*' and taginít na '*it is summer already*' as belonging to the non-predicative sentence of the impersonal-anaphoric type. (1917:152). It simply is not true that umúulán or taginít na are anaphoric, because these expressions can have no subjects and as such there are no subjects that can be deleted.

Bloomfield's lead on the analysis of Tagalog sentence types has been followed by linguists down to our own day. Schachter and Otnes (1972:546-547), for example, classify the sentences expressing natural phenomena like *umúlán* 'rain' and *umáraw* 'be sunny' as "derived sentences that occur as predicates of topicless sentences". (Ramos (1974:107), using Fillmore's case grammar model, says that "certain verbs that belong to the natural phenomena set take obligatorily suppressed instrumental force". She, therefore, analyses *umúlán* as the surface verb form of an original structure *umúlán (ang ulán)* 'rained (the rain)', in which *ang ulán* is 'suppressed'. McFarland (1976:2) calls *umúlán* 'it is raining' a "topicless sentence ... a statement which contains no sequence which also occurs as the topic of a basic sentence, and to which no such sequence can be added". Finally, De Guzman (1978:272), following the lexibase model, classifies verbs that refer to 'natural occurrences' as "phenomenal verbs ... a subclass of verbs (which) derives from norms marked [+ phen], and ... mean the occurrence of an event involving the natural phenomena designated by N. Unlike other verbs, they are marked for not co-occurring with a nominative actant".

All these descriptions recognise and explicitly state that there is a sentence type in Tagalog which does not have a 'topic/subject/nominative actant'. The implication is that the sentences of this type cannot be considered basic, since all basic sentences are bipolar, i.e. they have a subject and a predicate, and any sentence that lacks either can only be considered as 'anaphoric', 'derived', or a 'subclass'. While the grammatical model that posits this rule is adequate for most if not all the Indo-European languages, it certainly is not suitable for the great majority of Philippine languages.

To see whether this view was prevalent only with linguists who had worked on Tagalog but not with those who had worked on other Philippine languages, I tried to look at some of the more recent structural descriptions of a number of languages, and discovered that the same grammatical framework was employed in classifying sentence types. Thus, for example, Wolfenden's (1971) grammar of Hiligaynon also posits bipolar construction for the clause patterns, although he concedes that "the second pattern consists of only a predicate with or without modifiers of the headword. In these types of clauses, there is an obligatory absence of a topic". (1971:212). Similarly, Forman's analysis of Kapampangan sentences like *mumurán ya* 'it's raining', and *lílintík ya* 'it's showering' is that they belong together with greetings, curses and exclamations and should be classified as minor sentences "which do not have the bipartite subject-predicate construction of the major sentence types". (1971:126).

Outside the Philippine family of languages, the usual framework of grammatical description is likewise that of the bipolar subject-predicate or topic-comment structure. Thus, Soebardi (1973) fails to mention sentences like *hujan* 'it is raining', because it does not have this bipolar structure of the major sentence type; nor could it be classified under minor sentences, since it does not belong to the *ada*, the anaphoric response, to the vocative sentence types, nor to exclamations and greetings. Similarly, Walker (1976) divides all the sentences in Lampung into two types: major and minor sentences. "Major sentences have at least one topic constituent, which, however, may be deleted, and at least one comment constituent. Minor sentences have only one constituent, which functions as a comment, never as a topic". His major sentence types with topic deletion are anaphoric in nature (1976:12).

The contention of this paper is that this grammatical model which requires a bipolar or diadic structure of subject-predicate or topic-comment for all the basic structures of the language is not suitable for most, if not all, the

languages we have seen so far. The reason is that in these languages, there are monadic structures which are just as basic as the diadic sentences with subject and predicate.

It is, therefore, necessary to look for another grammatical system that can account for all the basic structures of this family of languages; a system that classifies sentence types according to their obligatory constituents – 'monadic', if it has one; 'diadic', if it has two; 'triadic', if it has three; and 'quadratic', if it has four. A model which requires bipolarity for all the basic sentences of a language simply will not be able to describe the languages mentioned above on their own terms.

STRUCTURE OF SOME AUSTRONESIAN LANGUAGES

To see if the grammatical framework described above can be used to account for the structures of other Austronesian languages, it was used to describe other Philippine languages, namely *Waray* (spoken on the island of Samar), *Bikol* (spoken on the southern part of the island of Luzon) and *Ibanag* (spoken in the Cagayan valley of northern Luzon). In addition, the model was also tried on *Bahasa Malaysia*, and the Indonesian languages of *Sasak* (on Lombok island, east of Bali) and *Acehnese* in Bandar Aceh in northern Sumatra). Finally, it was also employed in the description of the sentence types of Samoan, a language spoken on the island of Samoa in Polynesia.

The result in all instances was positive, and examples from the various Austronesian languages mentioned above are given in the Appendix of this paper. It may not be far-fetched to say that the model is perhaps applicable to many more, if not all, Austronesian languages.

The question now arises: what is the implication of all this to the comparative study of the Austronesian languages? Provided the model is valid and can adequately describe the structures of the Austronesian languages on their own terms, it should be a contribution towards facilitating the work of comparativists. As Blust (1974:2) sees it, syntactic comparative material can be organised in terms of the following scheme:

PERFECT comparisons:

1. identical
2. non-identical

IMPERFECT comparisons:

3. structurally non-identical, but all morphemes that occur are cognate
4. structurally identical, but only some morphemes are cognate
5. structurally identical, but no morphemes are cognate
6. structurally non-identical, and no morphemes are cognate

Imperfect comparisons pose problems, since a great number of inferences have to be made in justifying the reconstructions reached. On the other hand, fewer inferences need be made, if similarities (or even better identities) are found in the syntactic structures of the various languages being studied. It would, therefore, help if an identical framework is employed in the description of a group of languages belonging to the same family of languages, provided that such a framework adequately accounts for their structures.

APPENDIX

This section contains examples of the various sentence types of the following languages:

1. WARAY - a language spoken on the islands of Samar and Leyte in the Philippines - from Piczon-Llamzon 1973:88-89:
 - Quadratic* - pagqúulíq palaq ni huwán han líbru
'Juan has just returned the book.'
 - Triadic* - nagqíhaw an binatáq hin manúk
'The maid killed a chicken.'
nagpírit an irúy ña kumáqun an bátaq
'The mother insisted that the child eat.'
nagin (mahúsay ~ mádri) an darága
'The lady became (beautiful ~ a nun).'
 - Diadic* - nakáqun hi ándris
'Andres is eating.'
(prisidínti ~ buútan) hi Kennedy
'Kennedy is (president ~ good).'
 - hi huwán an sundálu
'Juan is the soldier.'
 - Monadic* - nalínug
'There is an earthquake.'
naurán
'It is raining.'
may(ada) ísdaq
'There is fish.'
2. BIKOL - a language spoken in southern sections of the island of Luzon - from Parma 1971:71-76:
 - Quadratic* - ipigbakál ni huwán an ákiq nin lápis
'Juan bought the pencil for the child.'
 - Triadic* - nagbakál nin siráq si laláki
'The man bought fish.'
puwídi sa laláki an kupiyáq
'The hat can possibly fit the man.'
kaipúhan kan ákiq an sapátus
'The child needs a pair of shoes.'

Diadic - malínig an salúg
'The floor is clean.'
duktúr si pídrú
'Pedro is a doctor.'
naghíhibíq an ákiq
'The child is crying.'

Monadic - nagtátagití
'it is showering.'
nagurán
'it rained.'
iguwá nin bisíta
'there is a visitor.'

3. IBANAG - a language spoken in the Cagayan valley in northern Luzon, Philippines - from Bauza 1972:77-81:

Quadratic - niyarádu nakketséru ta ummá in nuwáng
'The farmer used the carabao to plow the field.'

Triadic - pakanán ni huwán ikkítu
'Juan feeds the dog.'
iyáwaq namméstra illíbru
'The teacher will give the book.'

Diadic - mapiyá si dóming
'Doming is good.'
lumákag i abbíq
'The child walks.'

Monadic - magurán
'It will rain.'
nakkilikiláq
'The lightning flashed.'

4. SASAK - a language spoken on Lombok island, Indonesia - from fieldnotes with Mr Lalu Wiramaja of Mataram, Lombok, as informant:

Quadratic - beruq tiaṅ leman pokən
'I have been to the market.'

Triadic - áli ñəmbəlí mánok
'Ali killed a chicken.'
bápaq jári gúru
'Father has become a teacher.'
kúrsu ini ajína satus dólar
'This chair costs 100 dollars.'

Diadic - bápaq tínduq
'Father is asleep.'
bále bəlí
'The house is big.'
áli polísi
'Ali is a policeman.'

Monadic - újan
'It is raining.'
génter
'There is thunder'

5. BAHASA MALAYSIA - the national language of Malaysia.

Quadradic - kertas diperbuat daripada kayu
'Paper is made out of wood'

Triadic - ali pecahkan pasu
'Ali broke the vase.'
dia menjadi askar
'He became a soldier.'
ali ialah presiden
'Ali is the president.'

Diadic - rumah besar
'The house is big.'
ada orang
'There is a man.'
ali berlari
'Ali runs.'
itu ali
'That is ali.'

Monadic - banjir
'There is a flood'
hujan
'it is raining'

6. ACEHNESE - a language spoken in northern Sumatra, Indonesia - from Marhiah 1973, and fieldnotes with Halim Majid of Banda Aceh as informant:

Quadradic - nék gû pûnan čučo gûh fatimah
'Grandma names her granddaughter Fatimah.'

Triadic - si agam ji koh nalung
'The boy cut grass.'
gobnya kajût kû prèsidèn
'He became president.'
urûng jeh poh asè nyan
'The man killed that dog.'

Diadic - urûng sakét
'The man is sick.'
awakjeh ji plong
'They run.'
si din čarong
'Din is clever.'

Monadic - banjir
'There is a flood.'
 hujan
'It is raining.'
 khung
'There is a drought.'

7. SAMOAN - a language spoken in Samoa, a Polynesian island - from fieldnotes with Corporal Afamasaga of Dieppe Barracks, Sembawang, Singapore, as informant:

Quadratic - o lona tamā ua fa'anoa ina iā peta
'The father named him Peter.'

Triadic - tamā ua tapē le pua'a
'Father killed a pig.'
 o sione ole pelesetene
'John will become president.'

Diadic - fale telē
'The house is big.'
 ua moe sione
'John sleeps.'
 ole tamaloa leoleo
'The man is a policeman.'

Monadic - ua timu
'It is raining.'
 faititili
'There is thunder'
 uila
'There is lightning.'

BIBLIOGRAPHY

ANTTILA, Raimo

- 1972 *An introduction to historical and comparative linguistics*. New York: Macmillan.

BAUZA, Humberto G.

- 1972 A sketch grammar and lexicon of the Ibanag language. M.A. thesis, Ateneo de Manila University.

BLOOMFIELD, Leonard

- 1917 *Tagalog texts with grammatical analysis*. University of Illinois Studies in Language and Literature 3/2-4. Urbana, Illinois.

- 1933 *Language*. London: George Allen and Unwin; New York: Henry Holt.

BLUST, Robert A.

- 1974 Proto-Austronesian syntax: the first step. *Oceanic Linguistics* 13:1-15.

GUZMAN, Videa P. De

- 1978 *Syntactic derivation of Tagalog verbs*. Honolulu: The University Press of Hawaii.

LLAMZON, Teodoro A.

- 1976 *Modern Tagalog: a functional-structural description*. *Janua Linguarum Series Practica* 122. The Hague: Mouton.

McFARLAND, Curtis D.

- 1976 *A provisional classification of Tagalog verbs*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

MARHIAH, Syarifah

- 1973 A contrastive study of Acehnese and English syntax. Fakultas Keguruan Universitas Syiah Kuala, Darussalam, Banda Aceh.

PARMA, Marcial T.

- 1971 Basic grammatical structures and lexical forms of Bikol. M.A. thesis, Ateneo de Manila University.

PICZON-LLAMZON, Petrona

- 1973 Elements of modern Waray. M.A. thesis, Ateneo de Manila University.

RAMOS, Teresita V.

- 1974 *The case system of Tagalog verbs*. *PL*, B-27.

SCHACHTER, Paul and Fe T. OTANES

1972 *Tagalog reference grammar*. Berkeley: University of California Press.

SOEBARDI, S.

1973 *Learn Bahasa Indonesia: pattern approach*. 3 vols. Yogyakarta: Kanisius-Bhratara.

WALKER, Dale F.

1976 *A grammar of the Lampung language: the Pesisir dialect of Way Lima*. Jakarta: Badan Penyelenggara Seri NUSA.

WOLFENDEN, Elmer P.

1971 *Hiligaynon reference grammar*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

SUBGROUPS IN AUSTRONESIAN

Mark Harvey

0. NOTE

This is a considerably revised version of the paper presented to the Third International Conference on Austronesian Linguistics. The elaboration of some points owes much to the comments and criticism of such people as Robert Blust, Lawrence Reid, Raleigh Ferrell and David Zorc, to whom must go my gratitude. Especial thanks are due to William Foley for his valued remarks. I hasten to add, of course, that nonetheless any faults remain my own.

1. THE PAN PHONOLOGICAL SYSTEM

The PAN¹ phonological system has been and continues to be the subject of dispute in certain of its areas. The vowel system and a large part of the consonant system have been reconstructed with a reasonable degree of certainty. The phonemes which are generally accepted are:

- a) Vowels
- | | |
|---|---|
| i | u |
| ə | |
| a | |
- b) Consonants
- | | | | | | |
|---|------|----|----|---|---|
| p | t | ts | t' | k | q |
| b | d | dz | | | |
| | s | | | ɣ | h |
| m | n | | | ŋ | |
| | l, ɻ | | | | |

ts = t₂ (Dahl) C(Dyen)

dz = d₂ (Dahl)

t' = s (Dyen)

s = S (Dyen)

However, when one attempts to reconcile all the apical and palatal reflexes in the various languages, immense problems arise. This is the area where the

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 47-99. *Pacific Linguistics*, C-75, 1982.

© Mark Harvey

reconstructions vary greatly and where the dispute has been strongest. If one follows the usual principles of historical linguistics by positing a proto-phoneme for every set of correspondences which cannot be collapsed or explained away as being due to metathesis, dissimilation, assimilation, dialect borrowing, analogy or the myriad of other exceptions to the Neogrammarian hypothesis of exceptionless sound change, one still ends up with a system which has vastly more phonemes than any AN language. An example of this is the system Tsuchida (1976) uses in his reconstruction of Proto-Tsouic phonology. The phoneme inventory he uses is i, ə, a, u, p, b, m, t, T, C, j, n, ñ, k, g, ŋ, l, R, S₁, S₂, S₃, S₄, S₅, S₆, x₁, x₂, X, Q₁, Q₂, y, s, θ₁, H₁, H₂, N, L, w₁, w₂, W₁, W₂, C₁, t₁, D₁, D₂, D₃, D₄, z₁, j₁, ?₁, ?₂, ?₃, ?₄, K₁, K₂, l₁, rd, q, c, R₁, R₂, R₃, R₄, R₅. Tsuchida himself says that this number of proto-phonemes seems excessive and he states that he hopes that some of them, especially the ones separated by subscripts, will be ultimately unifiable. However, within the approach Tsuchida is using, one is forced to posit these proto-phonemes (and probably on examining more languages, a great many more). I would argue that the problems which have arisen in attempting to reconstruct the PAN phonological system have arisen not because of problems in the data but because of fundamental problems in the approach which has been taken to phonological reconstruction.

The Neogrammarian (hereafter N.H.) hypothesis of exceptionless sound change has never been observed as occurring in reality. As long ago as 1905 Gauchat (1905), in his study of sound change in Charmey, a village in the Suisse Romande, showed that sound change does not occur in the manner posited by the N.H., which states that sound change occurs without the knowledge of the speakers of a language as the result of imperceptible minute variations in the acoustic production of speech sounds.

In his study of palatalisation of *l* to *j* in Charmey, Gauchat found that generation I (the oldest) always had *l*^j and generation III always had *j* but generation II did not have some intermediate phonetic stage as the N.H. would predict. They had instead both *l*^j and *j* in fluctuation.

Labov, who has been the leading force in observing sound change in progress in recent times, has recorded essentially the same phenomena in his study of the centralisation of the diphthong *aw* in Martha's Vineyard (Labov 1963). Labov has shown in other studies of sound change in progress that not only does age affect sound change but other social factors such as class (his study of the reintroduction of *r* into the New York dialect: Labov 1972) and sex (his study of the raising of *æh* in New York: Labov 1966) also affect sound change. Studies of change in other English dialects have confirmed the fact that sound change progresses at different rates according to these social factors of age, class and sex (see Trudgill's study of vowels in Norwich English: Trudgill 1971).

However, this does not mean that sound change occurs at random. Both Labov and Gauchat found that sound change was favoured by particular environments. Furthermore, over time the usage of the variants tends to become more and more consistent. Either one variant will win out over the other or they will be used in different environments. (e.g. Gauchat's generation III speakers always use *j* and Labov's generation III on Martha's Vineyard have a consistent environment for the centralisation of *aw*).

Other factors may interfere with this development towards consistency. Labov found in his study of the raising of *æ* in New York that this pronunciation was a stigmatised one and that people who raised the *æ* would sporadically correct it in situations where they were being careful in their speech. It is well known that the prestige of the dialect, class or person who makes an innovation in the

phonological system has great influence on how far it spreads. Furthermore, even though there is a strong tendency for consistency to develop, it does not always do so. In his analysis of the development of Middle English a:, ε: and e: into Modern English (Labov 1974), Labov regards the development of 'steak' as irregular. In the rotacisation of medial s to r in Latin the retention of s in nasus is irregular (Anttila 1972). These irregularities are not due to social or other types of interference but simply to the failure of the sound change to become consistent in all words.

Therefore, it appears that the N.H. does not correspond to reality and it fails to predict the development of sound change when it occurs. The N.H. is not only unsound in its relationship to observed reality of sound change, it is also unsound theoretically. The clearest and most incisive attack on the theoretical standing of the N.H. was made by Hugo Schuchardt in his essay "On sound laws: against the Neogrammarians" (Schuchardt 1885). Schuchardt's attack is not a purely negative one: he presents a clear picture of how he views phonological change occurring.

Schuchardt wrote his essay in pre-phonemic times and it must therefore be reconsidered carefully in the light of phonemic theory. The same thing applies to the N.H. as well. Indeed phonemic theory raises immediate problems for the N.H. because phonemes can only change abruptly. A particular sound is either an allophone of one particular phoneme or it is not, but there is no half-way house. It has always been known that certain types of sound change such as metathesis must be abrupt and these types of sound change have always been problematic for the N.H. We will consider the question of the gradualness of sound change later. For the moment we will return to Schuchardt's criticism of the N.H.

The first, most important and probably the most disputed point raised by Schuchardt is that a phonological system is a psychological entity and as such it should not be affected by purely physical factors such as the continual acoustic variations which occur in speech. Labov demonstrates in his article "On the use of the present to explain the past" (Labov 1974) that speakers may consistently differentiate classes of sounds that they regard as one phoneme. As Labov puts it "small differences do exist. To the native speaker they may be worth noticing and may in fact be impossible for him to label; nevertheless, he regularly produces these differences". If phonological systems can ignore differences which are consistently made, why should a random acoustic variation cause it to alter? Schuchardt argues that sound change is therefore as psychologically conditioned as analogy and borrowing are (i.e. that an alteration in a psychological system must be made by psychologically controlled factors). This point of course depends on accepting Sapir's (1949a) view of the phoneme as something real in the mind of the speaker and not Twaddell's distributional definition (Twaddell 1935). What Schuchardt is saying essentially is that while speech varies continually, this in itself does not cause alteration to phonological systems. For this to occur the variation must be pushed by the same things which control the phonological system.

In furtherance of this argument Schuchardt points out that sound change and analogy interact on one another. If they were motivated by entirely different forces, sound change by physical factors and analogy by mental ones, then they would not affect one another because they would be operating on entirely different planes. From this it follows that sound change and analogy have the same motivation by mental factors (as nobody would wish to argue that analogy is motivated by physical factors).

As a part of his general consideration of the relationship between analogy and sound change, Schuchardt attacks the view that analogy causes irregular and arbitrary sound change. His own example of analogy causing regular sound change is the change of all Old Spanish and Portuguese participles ending in *-udo* to *-ido*. Schuchardt here is arguing for grammatical conditioning of sound change, a phenomenon which now has wide acceptance (see Anttila 1972).

Schuchardt's final point on analogy and sound change is that there is no clear dividing line between them. He gives two examples of the development of Italian from Latin to show this. Latin *comite* becomes Italian *conte* (it should be *comte*) and the Latin phrase *non gravis magis leve* becomes Italian *non greve ma lieve* (it should be *non grave ma lieve*). These are both examples of assimilations though the motivations are different. In the first case the motivation is to reduce the number of places of articulation and in the second it is the association of the words in that particular phrase. Yet both are assimilations, changes that make things more like one another for whatever reason. While it is clear that there are differences between the two assimilations, it is equally clear that there are similarities which would be ignored by saying that one is sound change and the other analogy and never the twain shall meet.

If, as Schuchardt argues, sound change and analogy are motivated by the same factors, then sound change will like analogy not proceed to produce consistent results. This, in fact, is what we find in reality. It has long been known that despite what the N.H. said should happen, sound change does not always proceed regularly.

Workers in lexical diffusion theory have provided extensive data confirming this and they have attempted to provide explanations for sound change as they have observed it. The theory of lexical diffusion is discussed in Wang (1969) and further discussed and exemplified by Chen and Wang (1975) and Krishnamurti (1978). Lexical diffusionists take the view that sound change is phonetically abrupt and lexically gradual (Wang 1969:14). In other words sound change spreads gradually across the lexicon and words may have more than one pronunciation. This means that words which were at one time homophonous can acquire different pronunciations over time (Chen and Wang 1975:261). Furthermore, while sound changes tend to apply to all relevant lexical items over time, they can peter out or meet a competing change.

Chen and Wang (1975) and Krishnamurti (1978) provide extensive data which clearly shows that sound change does proceed through the lexicon gradually. Neither Chen and Wang nor Krishnamurti address themselves to the causality of sound change. Chen and Wang argue that the motivations for particular phonemes altering in particular ways are the articulatory and perceptual constraints of speakers (i.e. sound changes occur in particular ways for physical reasons).

When one considers the causality of sound change within a lexical diffusionist view, it is obvious that sound change cannot be caused by acoustic/physical factors. If one accepts that sound change and analogy are motivated by the same factors, then it is comprehensible that sound change can spread gradually. For example, if a language has a number of suffixes marking the dative case and one of these affixes is analogised to replace the others, it will take some time for this process to be completed and it may never reach completion.

In his 1969 article Wang quotes Sturtevant (1917:82) with approval "We have seen that many sound changes are irregular when they first appear and gradually become more and more regular. The reason is that each person who substitutes the new sound for the old in his own pronunciation tends to carry it into new words.

The two processes of spread from word to word and spread from speaker to speaker progress side by side until the new sound has extended to all the words of the language which contained the old sound in the same surroundings". Sturtevant said, in addition "Such a spread of a sound change from word to word closely resembles analogical change; the chief difference is that in analogical change, the association groups are based upon meaning, while in this case the groups are based upon form". (1917:80)

This is exactly what Schuchardt said in his discussion of the relationship between analogy and sound change. Wang points out that from observation, the sound changes that occur are few in kind. Given this fact and the fact that a sound change spreads through the lexicon over time, there is a very good chance that two or more sound changes may operate on the same material to produce irregular results because the rules will operate on different items at different times and therefore produce different results.

Sturtevant essentially agrees with the views of the lexical diffusionists discussed earlier. It is clear therefore that sound change can produce irregular results.

It is not only the lexical spread of sound change which presents problems for the N.H. The geographic spread of sound change has always been exceedingly problematic. The N.H. states that a particular sound change occurs without exceptions but it does so only within a uniform speech community over a certain period of time. This uniform speech community is the 'same dialect' of the Neogrammarians.

However, no such uniform speech community has ever been found, nor is there any reason to imagine, given religion, politics and economics, that any such community has or would ever exist. To get this community Delbrück (1884) had to define it as the normative average of the speech of an individual at one point in time. Quite apart from the fact that this definition ignores things such as individuals controlling social registers, the difference between a regular change in this community and sporadic phonological change of a group of people is for all practical purposes minimal. Furthermore, as Schuchardt points out, even the speech of individuals is subject to variation (i.e. register control, the difference between speech to strangers and intimates). This discussion should not be taken to deny the existence of speech communities, which are as Labov says (1974) orderedly heterogeneous, only the existence of *uniform* speech communities.

This is not the only area where the N.H. fails to correspond to reality. The N.H. claims that sound change proceeds without any conscious knowledge of the speakers. Metathesis, which can sometimes be stated in the same way as any 'ordinary' sound change, has always been problematic for this viewpoint. In Slavic, for example, all sequences of (C)VRC go to (C)RVC (where R represents r or l). Furthermore, even such Neogrammarians as Delbrück (1884) state that certain sound changes spread because they are considered fashionable or prestigious. This would appear to be irreconcilable with the view that sound change is unconscious. Also, as Vennemann (1972) states, people in any speech community are aware of differences between their speech at various points in time, and of the differences based on all the parameters that Labov has studied (Vennemann cites the change of *hw* to *kw* and *xw* in 20 years in Icelandic).

After having raised these points against the N.H., Schuchardt goes on to present his own view of how sound change takes place and regards sound change as occurring in quite the opposite direction from the Neogrammarians. All change in phonological systems occurs initially as a sporadic change. The only difference between 'regular' and 'irregular' sound change is that 'regular' sound

change is 'irregular' sound change which has been generalised to consistency. Schuchardt points out that it is frequently possible to trace the spread of a phonological change from some area of origin as it spreads to more and more speakers. Vennemann (1972) gives an example of this and also of the generalisation of a sound change in the development of $s \rightarrow \int$ which occurs as follows:

$s \rightarrow \int$ /l Norwegian
 $s \rightarrow \int$ /lrmnw North German
 $s \rightarrow \int$ /lrmnwpt Standard German.

Schuchardt's own example concerns the development of \bar{e} and \bar{o} to ie and uo in Italian. He shows that it started out when there was a following i and u respectively. Some dialects still have it in only this environment, but others have spread it in various stages. In some dialects there are only forms such as *vieni*, *buonu*, *buoni* and the like. However, other dialects spread the changes through paradigms so you get forms like *viene* and *buona*. Still other dialects have forms such as *pietra* and *ruota* where the original conditioning factor is completely lost.

Schuchardt points out that his view of sound change is closer to how sound change occurs in reality than is the N.H. which is essentially a philologist's view of how sound change occurs. This is not to say that there are no problems with the alternative view of sound change Schuchardt is proposing. The first is that he gives no example of a truly sporadic change which generalises. All the changes which he discusses as generalising have some initial phonetic or grammatical conditioning. This is in agreement with what Labov and Gauchat found in their studies. Indeed Labov takes great care to point out that while in an overall speech community a sound change may proceed with great irregularity and produce irregular results, if you analyse the speech community along the social parameters Labov uses you will find sound change proceeding with a great deal more regularity, along phonetic and grammatical parameters, within the social parameters.

Therefore it seems that while Schuchardt's view of sound change is closer to reality than the N.H., it does not correspond exactly to reality. Therefore, I think that you would have to modify Schuchardt's statement that all sound change is in origin sporadic to correspond with Labov and Gauchat's observations. The restatement would run something like 'sound change is not in origin consistent but tends to occur more in certain environments than others'.

The second problem with Schuchardt's view is that it cannot explain things such as Grimm's Law where classes of phoneme alter and follow one another in a chain. The N.H. offers no explanation for these phenomena either. Explanations of these phenomena have, I think, been offered and generally accepted and these explanations can be incorporated into the view of sound change espoused by Schuchardt and Labov. The first which explains why classes of phonemes act together as a class, is the concept of features in phonology. There is no reason why any view of sound change should be restricted to dealing with phonemes only. Schuchardt's view is certainly not so restricted.

Martinet's (1952) concept of push and drag chains, whereby when one phoneme or class of phoneme alters, the nearest one will move into the range vacated, is a generally accepted explanation of chain sound changes. It may be added to Schuchardt's viewpoint without any problems.

The third and most major problem with Schuchardt's and with Labov's views and observations of sound change is to explain why sound change generalises.

There is no doubt from Labov and Gauchat's observations that even though sound change may initially be irregular, it does become regular by generalisation.

Andersen (1973) in his article on "Abductive and deductive change" proposes explanations for the genesis and development of sound change. The first type of change Andersen discusses is abductive change. According to Andersen, abductive change occurs when children are presented with phonological data and they analyse it in a different way from their parents. This means that they have different phonological categories from their parents. Andersen goes on to claim that in order to approximate their speech to that of their parents they have rules which produce surface forms corresponding to the accepted norm. These rules are called A rules. Andersen states that these rules may be elaborated and revised throughout the speaker's life.

The second type of change Andersen proposes is deductive change. This type of change involves altering the speaker's surface forms to correspond to the underlying phonological representations. These changes occur abruptly in each word, but across the vocabulary they occur gradually depending on social factors. Old forms will tend to be used with the older generations.

These two types of sound change would certainly go a long way to explaining why sound change generalises. If a reanalysis of the phonological system is involved, then one would expect all instances of the phoneme(s) affected to be altered. However, this is not necessarily the case because as we know that sound change does not always affect all instances of a phoneme (see Labov's (1974) discussion of the Middle English development of e: ε : and a:). Furthermore, in his study of his child's acquisition of the English phonological system and its categories, Smith (1973) found that while his son normally produced words which corresponded to the correct English phonological form in terms of the child's own phonological system, there were occasional exceptions when the child had miscategorised something. Therefore, it does not seem that, if children did categorise differently from their parents, they would necessarily do this without exceptions.

Another part of Andersen's hypothesis that would need some testing before being accepted is the A rules. Labov in his study of centralisation does not mention whether the middle generation of speakers in Martha's Vineyard tended not to centralise their vowels when talking to the older speakers. Labov does discuss how New York speakers tended to alter their raised vowels when they were speaking very carefully. They did this sporadically, however, and they did not do it in ordinary speech. This could also be regarded as an example of Andersen's adaptive change (changes due to factors outside the linguistic system). Abductive and deductive change are types of evolutive change (a change explicable in terms of the linguistic system which gave rise to it).

Adaptive change occurs when the native forms of a particular area are replaced as the result of extralinguistic pressure. This would most often involve the replacement of items or pronunciations of less prestigious communities by those of more prestigious communities. This is essentially the same phenomena as Andersen's A rules. These two are random assimilations of less prestigious (children's) speech to that of the more prestigious (parents'). Both adaptive change and A rules are simply examples of socially motivated alterations in a speaker's pronunciation. It is quite conceivable that a particular variant may not be socially unacceptable and so A rules may as such never operate. We have already seen in Wang (1969:15) that words can have variant pronunciations. In the examples he quotes neither variant appears to be socially stigmatised.

Andersen's hypothesis would certainly appear to provide a good explanation for the linguistic behaviour of the youngest generations in Labov's and Gauchat's studies. These were the ones who had the sound changes consistently in particular environments. However, it cannot deal with the linguistic behaviour of the middle generation who were not consistent in their use of changed forms. Until the necessity of A rules for all cases of sound change can be proved, one is entitled to assume there are problems with Andersen's theory in relation to the observed behaviour of the middle generations in Labov's and Gauchat's studies.

Perhaps the most important point about Andersen's view of sound change is that he, like Schuchardt, posits a psychological motivation for sound change. While it may not be able to explain all the observed facets of sound change, it does seem to provide a good explanation for the final stage of sound change as observed by Labov and Gauchat.

As Andersen points out, if you take a view of sound change something like his, then the question of whether sound change is gradual or abrupt loses much of its meaning. Sound change, when it involves abductive change, is neither gradual nor abrupt; deductive and abductive change are phonemically abrupt but across the vocabulary they spread gradually. Basically, according to Andersen, when it is meaningful to talk of gradual vs abrupt, sound change is psychologically abrupt but its linguistic manifestation is gradual.

However, I would argue that unpredictable correspondence sets are a natural consequence of sound change. Wang, Labov and Andersen agree that while sound change starts out being irregular (though associated with conditioning either phonological or grammatical), it will over time produce consistent correspondence sets. However, as Labov demonstrates, this progression towards regularity holds only within particular groups in the speech community. The development of a sound change when the speech community is looked at as a whole can be quite irregular. Once the various component groups of the speech community start to influence one another's speech, as they are continually doing, then a combination of different conditioning factors and different stages of development can cause irregular results. As Schuchardt points out, it is meaningless to call this dialect borrowing. Within a speech community a sound change must originate somewhere and from there it spreads. The social boundaries which are so important in determining how a sound change develops are not absolutely rigid. Therefore, it is natural that the development of a sound change in one section of the community will influence and be influenced by its development in all sections of the community. As the geographical boundaries of speech communities are not rigid either, it makes very little sense to attribute inconsistencies in the result of one sound change to 'borrowing' and then to describe another sound change with consistent results as having spread throughout the same area. The term borrowing should be reserved for those cases where you have languages which are not mutually intelligible or which are geographically separated (i.e. in those cases where you cannot have a common speech community). Borrowing within a speech community is something that is entirely to be expected.

This is essentially what Bloomfield (1933:364) says:

We must suppose that, no matter how minute or accurate our observation, we should always find deviant forms, because from the very outset of a sound change, and during its entire course, and after it is over, the forms of the language are subject to the incessant working of other factors of change, such as, especially, borrowing and analogic combination of new complex forms.

From this discussion of sound change it can be seen that while sound change has a tendency over time to produce consistent results, a number of factors may in the natural course of development of any sound change operate to prevent the sound change producing consistent results.

The view of phonological change, which we have presented, naturally affects the method we use to reconstruct an inventory of proto-phonemes. Because we accept that change can and does produce inconsistent results, we cannot accept all correspondence sets as valid. There is no guarantee that all correspondence sets which are different and which cannot be shown to be conditioned variants do in fact represent different proto-phonemes. Therefore, we will need other factors to determine the acceptability of correspondence sets as evidence for particular proto-phonemes.

Before accepting a correspondence set as evidence for a proto-phoneme we will require that it should occur in a group of cognates which occur throughout Austronesia. If a correspondence set occurs in a set of words which occurs only in one area and not others, then it cannot be accepted as PAN because it could easily have been innovated in that area. We will also require that it should occur in more than one or two words because a correspondence set involving only one or two words could arise by sporadic phonological change.

Before accepting a correspondence set, which occurs only in a small number of non-basic items of vocabulary, as representing a separate proto-phoneme, we will require that the correspondence set have reflexes in a number of geographically separated languages which serve to separate it from other correspondence sets. Otherwise we could again be dealing with an areal innovation.

Taking the PAN phonological system as set up by Dahl (1976) as a starting point, and taking Dempwolff (1934-38) into consideration also, we will consider what changes should be made in the light of the criteria proposed above. The system Dahl sets out (1976:101) is:

Vowels: i u
 ə
 a

*i and *u may be syllabic or non-syllabic.

Consonants: Stops and affricates Nasals Continuants

	Stops and affricates		Nasals	Continuants	
	Voiceless	Voiced	Voiced	Voiceless	Voiced
Bilabials	p	b d ₁	m		
	t ₁		n(1)	ʈ	l(1)
Apicals		d ₂		s	r
		d ₃ d'			
Palatals	k'	ʒ g'	n'		
	t'				
Velars	k	g	ŋ		ɣ
Uvular	q				
Laryngeal				H?	

We agree with what Dahl says (1976:55-69) concerning Dempwolff's *t̥ and *d̥. Dahl says these two phonemes were innovated under Indic influence, in the same way the voiced aspirates were innovated in Madurese. This led to their cognate sets being confused and contradictory. We also agree with the suggestion that the languages where t̥ and d̥ are found originally had a dental t and an alveolar/retroflex d, as is found in some Western Austronesian languages still (Henderson 1965). This situation results from the merger of an original *d and *d̥ (quite separate from Dempwolff's *d̥), and made the languages especially receptive to the introduction of phonemic retroflexes from Indic.

We will now go on to consider Wolff's (1974) proposal to remove *r from the PAN phoneme inventory. He argues that all correspondence sets which are cited for *r can be dealt with as reflexes of either *d or *ɣ. In the Oceanic languages *d and *r are always reflexed as r. Indeed, according to Wolff it is only TB and Ml which require us to establish *r as a separate phoneme. Furthermore, Wolff states that he can find only two cognates for r in Taiwan. However, Tsuchida (1976:177, 227) provides several cognates for *r on Taiwan, five of which I find convincing, and which, I think, compel us to reconstruct *r. The cognates are:

1. *garuts Pw garuts 'comb', Ml garut 'scrape', Sa keru 'to scratch'.
2. *kurap Sar kurapə 'scabies', Pw kurap 'a scaly skin disease', Ml kurap 'ringworm', Tausug kulap 'scabies'.
3. *buar Kn ma:-vu:-vuare 'to dissemble after a meeting', Sar m-i:-vuare 'to collapse, tumble down on piles of firewood, stone', Tg buwal 'fall, fell, demolish, drop, fall on ground'.
4. *kuruŋ Pw kuruŋ 'wooden frame for carrying firewood', Ts t-h-i-kruŋu 'surrounding walls of a pig pen', Ts re-kruŋv-a no hopo 'mosquito net', Tg kulón, Ilokano kurún, Bikol kurún 'fortress', Tausug ku:ŋan 'cage', To kolo 'fortress, temporary fence of native cloth around a grave', Sm ʔolo 'fort'.
5. *gərgər Pw m-igərgər, Kan m-a:-kərəkər, Sar k-um-a-kə:rə kə:rə, Ts s-m-o-ʔərʔərə, Ilokano pi-gərgər, Kampampangan galgal, Proto-Minahasan *gərgər Ml gəgar 'to tremble'.

In Paiwan, Saaroa, Kananabau none of the other voiced apicals *d, *l or *d̥ ever has an r reflex. In Tsou *d̥ gives r but it is the only voiced apical to do so. In all the Tsouic languages *ɣ gives r; in Paiwan *ɣ gives ø normally but r occasionally. In Tg, Bikol and Tausug *ɣ gives g. In Kampampangan *ɣ gives y.

Taking the cognates one at a time, we will show that *r is the proto-phoneme to be reconstructed rather than *ɣ or a voiced apical. For the first cognate we must note that the Paiwan evidence shows that this is derived from gutsguts 'to scratch' and the infix ar (having sound or quality of; involving the use of). Now this -ar- in Paiwan could come from *-aɣ- or *-ar-. However in Tg these two are reflected as -ag- and -al- being basically verbalising and adjectivalising infixes respectively (Zorc, personal communication). The Paiwan -ar- is obviously cognate to -al- not -ag-. Therefore we have to reconstruct *-ar-. An infix with *r is extremely strong evidence for the phoneme's existence.

The second cognate is not quite so convincing: the Pw, Ml and Sar forms could come from *kuɣap but the Tausug form cannot. The Tausug form could be a Malay borrowing as it is the only Philippine cognate. Therefore *kuɣap is not too certain. The third cognate is much stronger: the Kn and Sar forms cannot

reflect a voiced apical and Tg cannot reflect *ɣ and there is no Malay form so Tg cannot be a borrowing. The fourth cognate is equally strong. The Paiwan form excludes a voiced apical. The Tg, Bikol and Tausug forms exclude *ɣ. The Philippine forms do not appear to be borrowings. The Tg, Bikol and Ilokano forms all have final accent whereas if the borrowing was from Malay one would expect at least some forms with initial accent (Zorc, personal communication). The Tausug form shows the loss of l (*r→l in Tausug) between similar vowels which is at least 1,000 years old (Zorc, personal communication). The fifth form provides perhaps the strongest Philippine evidence. Pw, Kn, Sar and Ilokano all exclude a voiced apical (r cannot be a reflex of a PPh voiced apical in this position (Zorc, personal communication). Kampampangan excludes *ɣ. The Philippine forms cannot be borrowed from Malay as the Malay reduction of *gəgər to *gəgər is very ancient.

Therefore it appears that we must reconstruct *r. However, we should note that *r does not occur initially. This is significant in light of the fact that a number of Austronesian languages have r as the intervocalic allophone of d. The difficulty of finding cognates for *r and its defective distribution both suggest that it was a very rare phoneme of recent origin in PAN.

Though most proposals for alteration of the system originally proposed by Dempwolff have involved reducing the number of phonemes, there has been one set of proposals to increase the number of phonemes which is generally accepted. These proposals arise because of Taiwanese data which was not available to Dempwolff. The new phonemes are *ts, *d₂, *ɖ (different from Dempwolff's *ɖ) and *ɣ. Outside Taiwan *ts merges with *t, *d₂ and *ɖ with *d, and *ɣ with *n and *l. The correspondence sets for these phonemes involve such basic and widespread words as 'eye', 'excrement', 'ear', 'die', 'water', 'black', 'two', 'hear', 'child', 'swim' (see Dahl 1976:55-59).

The alternative to accepting these new proto-phonemes is to regard them as unconditioned areal splits (*ts being *t, *d₂ and *ɖ from *d and *ɣ from *l and *n). Each of these possible splits will be considered. The idea of initial ɣ being the result of an unconditioned split from *l and non-initial *ɣ being an unconditioned split from *n completely lacks credibility; if indeed it is possible that such a sound change could occur. Therefore *ɣ must be accepted.

The idea of a double split of *ɖ and *d₂ from *d also lacks credibility. We will consider *ɖ first. In Amis, Tsou, Thao, Favorlang, Paiwan and Siraya *ɖ is reflexed separately from *d (see Dahl 1976:64). In Amis and Thao *ɖ is reflexed as t and *d as r and ø respectively. In Tsou, Favorlang and Siraya it is r whereas *d is ts, t and d respectively. Too much reliance should not be placed on Favorlang and Siraya owing to the paucity of cognates and uncertainties as information is available only from 17th century Dutch manuscripts. Nevertheless, it does appear from Marsh's (1977:76-79) study of Favorlang that *d and *ɖ have different reflexes. The reflexes of *d and *ɖ are sufficiently different and consistent for us to say that there appears to be enough evidence for *ɖ. Also as Dahl points out the existence of a PAN *ɖ would help to explain the fact that t is dental and d is alveolar in certain western AN languages (Henderson 1965). We would have a merger of *d and *ɖ to ɖ while *t would remain dental.

The evidence for *d₂ is very much weaker than that for *ɖ. The main evidence comes from Paiwan (see Dahl 1976:59), with support to some extent from Puyuma, Favorlang, Siraya and Kavalan. Of these the Puyuma evidence appears to be unreliable, and even if it is reliable, the reflexes in Puyuma of the voiced apicals are too varied to form a clear basis for separating *d₁ and *d₂.

Favorlang appears to have merged *d₂ and *ɖ (see Marsh 1977:59-61, 76-79). However, as we pointed out earlier the material available on Favorlang is not reliable enough to be a base for positing a distinction in PAN of *d₁ vs *d₂. With Kualan and Siraya there are simply not enough cognates to support the distinction for PAN. Siraya has the following *d₂ cognates salom *d₂aɬum 'water', ricos *likud₂ 'back' and houkas *t'ukud₂ 'stick' [found only in Taiwan]. Siraya has d for *d₁. Furthermore Siraya has the same problems as Favorlang. The only sources we have on the language are 17th century Dutch sources which are, as we have said, not a reliable enough base for a PAN proto-phoneme distinction. Kualan has only one *d₂ cognate ranum *d₂aɬum 'water'.

This is not sufficient evidence to discount the possibility of an areal innovation, given that the reflexes of *d₁ and *d₂ are phonetically similar. There are variations in the reflexes of *d₁ in several Taiwanese languages (Dahl 1976:64). Therefore unless more evidence becomes available from a more detailed study of the Taiwanese languages, we must regard the case for *d₂ as unproven.

On the other hand, *ts is reflexed separately from *t by all the Taiwanese languages except Amis, Bunun, Siraya and Kualan. Furthermore, there are extensive cognate sets to attest the difference (Dahl 1976:61-64). Furthermore, *ts is fairly consistently reflexed. If *ts was an unconditioned areal split from *t, we would expect there to be variation among the languages (e.g. *matsa being reflected as *mata in some languages which were being used as criteria for separating *ts from *t). Therefore, the evidence appears good enough to accept *ts.

We have now dealt with the apicals and we will now go on to consider the palatals. This is the most problematic area in reconstructing the PAN phonological system; the area where the correspondence sets are least consistent. The first set of palatal phonemes we will consider are Dempwolff's *d' and Dyen's *Z (Dyen 1951). These two phonemes are obviously very similar in nature as they have the same reflexes with different distributions.

The two languages which require us to reconstruct a *d' separate from *Z are Javanese and Toba Batak. The Javanese and Toba Batak reflex of *d' is d' and of *Z is d, but M1 shows d' for both. However, if we examine the cognates we find that in Javanese and Toba Batak d is the reflex you get in basic words such as 'road' *Zalan, 'rain' *quZaɬ (two of the most widespread cognates throughout Austronesia), 'tongue' *Zilaq, 'compass-needle' *Zayum (the Austro-nesians were and are a seafaring people). In Javanese there are at least two cognate sets which have inherited Javanese words with d and Malay loans with d'. There are d'arum 'mediator' (must be from Malay because *ɣ gives ø in Jv and r in Malay) vs dom 'needle' (Native Javanese); and d'uruh 'syrup' (must be a Malay loan for the same reasons as before) vs du-doh 'sap'. In his discussion of *Z Dahl (1976) discards two Toba Batak words pod'am 'close the eyes' and in'd'am 'borrow' (this has a cognate in Paiwan) as Malay loans. We know from Dyen's (1956) work on Ngadju Dayak that Malay can have an immense influence on the vocabulary of languages. In Ngadju Dayak the Malay loans form a greater percentage of the vocabulary than the inherited forms. In this case Dyen states that to determine the native reflex one must accept the reflexes given by the basic words. We know from inscriptions that Malay has been known in Java for 1,000 years and its influence has steadily increased through time. Furthermore, Madurese appears to have d' or d'h as its reflex of *Z. It appears therefore that there is good reason to say that Javanese and Toba Batak d' have come from Malay or Madurese loans. This would mean that there is no necessity for reconstructing *d' as a phoneme separate from *Z. There is another very good reason

not to reconstruct *d' as a phoneme separate from *Z. There are only two cognates for *d' on Formosa. One is a plant name (admittedly an important one, the betel-nut) and the other a word for 'sharp'.

Jv d'ambe Ml d'ambi 'areca palm', Md d'ambe 'a vegetable', Pw qa-zavai 'Machilus Kusanoi'.

Amis si-tarum, Ss s-əm-aəm, TB tad'om, Jv tad'əm, Ml tad'am, NgD tad'im 'sharp'.

The cognate for 'sharp' raises some problems for our argument that the Javanese and TB forms with d' are Malay loans. The Jv and TB forms have the native ə and o reflexes of *ə as opposed to the Ml a reflex. There are two possible answers to this problem. The first is that the form was borrowed from Malay when *ə was still ə in all three languages. The other is that the presence of d' in Jv and TB results from the spread of the phoneme once it was established in Jv and TB by Ml loans. This would be a type of hypercorrection. Exactly the same phenomenon occurred in Madurese with the voiced aspirates which were established by Indic loans (e.g. *damay Md dhamar). If Md can replace d with dh, it is perfectly possible that Jv and TB have replaced d by d'. I prefer this explanation to the first one proposed, because it correlates with what we know happened in Madurese and with what has happened elsewhere in the world, whereas there can be no proof that what is proposed by the first explanation ever happened.

Dahl and apparently Dempwolff (Dahl 1976) suspect that *d' was not an original PAN phoneme. Unlike most other PAN consonant phonemes it occurs only initially and medially. As we have examples of its spread now, we could reasonably suspect that it has spread since PAN times and that at one stage it was not a separate phoneme, though it certainly must be reconstructed as such for PAN. A possible explanation for the spread *d' is suggested by the fact that Malay has been in contact with the Austroasiatic languages, which have a full series of palatals in all positions of the word, for a long time. Since Emeneau's article on India as a linguistic area (1956) it has been well known that certain classes of phonemes can spread throughout an area without actually being borrowed. As Malay already had d', coming into contact with the Austroasiatic languages would provide an impetus for the spread of d'. Once it had spread in Malay, it could spread in other AN languages, because of the use of Malay as a trade language throughout Indonesia.

We will now go on to consider the status of *n' as a PAN phoneme. We know from Blust's (1978) work that this phoneme occurred in the ancestor language of the Oceanic languages. It occurs in a number of Western Indonesian languages such as Javanese. It occurs in Kapampangan and Chamorro and so must be reconstructed for PPh (Blust and Zorc, personal communication). However, in Formosa I have been able to find only three cognates with *n' from Tsuchida (1976:229). They are as follows:

1. Pw sə-qaʔud' 'to be carried away by current', Kn m-a-ʔatsunu, Sar m-u-aʔusa, To ŋ-ohtsu, Ruk mu-aʔuʔu, mu-luʔu, mu-ʔuʔu 'to be adrift', Bun maŋ-hanuʔ, Ss lö-ʔälur, Pl mu-lahuʔ 'flow', Tg ʔanod, Hl Bs anod, Ml haʔut, Jv aʔut, To mu-ʔanu 'float' (Am ma-qalul, Se qəlul, Ruk ʔarur 'point to a form' *qalul)
2. Pw qa-ʔits 'skin, pod shell', Kn ʔaŋi:tsi, Sar ʔaʔitsi, Ts hitsi 'leather', Kel anit 'skin, bark', Tg anit 'scalp', WBM anit 'leather', Bk anit 'animal hide', Tebeduh Piching aʔit 'skin'.
3. Kn ʔa:la 'flame', Thao ʃu-na:raʔ 'to burn', Tg nala 'flame', Sa nala 'to explode, as steam'.

As can be seen Kn is the only language with reflexes in all three forms. It has two separate reflexes; η and n , which are not conditioned in any obvious way. If we examine Tsuchida's material (1976:229-230) we see that Kn η corresponds to Sar n and \dagger (which in turn correspond to $*n$ and $*\dagger$). We know that in Kn $*\dagger$ and $*n$ have fallen together as n . Therefore it seems that the best explanation for the η in Kn, where we would expect n , is that n has sporadically changed to η in Kn. Therefore, we can regard the Kn reflex in all three cases as n . When we examine the other Formosan forms we see they point to forms $*qa\dagger ud$, $*qa\dagger its$ and $*nala$. In Thao the reflex of $*n$ is n and $*\dagger$ is \dagger or l ($*ini$ 'this', Th $ináy$). Therefore Thao $\int u-na:ra$ indicates a form $*nala$ rather than $*\dagger ala$. Paiwan reflects $*\dagger$ as \dagger and $*n$ as n ($*panaq$ 'arrow', Pw $panaq$, $*da\dagger um$, Pw $za\dagger um$ 'water'). Therefore Paiwan indicates $*qa\dagger ud$ and $*qa\dagger its$. Unfortunately we do not have a Paiwan reflex for $*nala$, nor do we have any from any of the other Formosan languages which distinguish $*n$ and $*\dagger$ — in fact all the languages except Kananabu, Bunun, Puyuma and Kavalan (Dahl 1976). So far we have been working on the basis of $*n'$ and $*\dagger$ merging to \dagger in Formosa. Now we are faced with the possibility that n' is reflected as $*n$ and $*\dagger$. While in our view of phonological change an unconditioned split of a phoneme is not impossible, other solutions to the data should be considered. It could be argued from the forms $*nala$, $*qa\dagger ud$ and $*qa\dagger its$ that $*n'$ has split to $*n$ initially and $*\dagger$ medially. Against this one may raise evidence from Paiwan that $*qa\dagger its$ consists of a prefix qa and a base $*\dagger its$ (Ferrell 1982). Therefore if $*n'$ merges with $*n$ initially we would expect the Formosan languages to reflect $*qanits$ (unless one maintains that the split merger took place after $*n'its$ ceased to be used on its own — in Paiwan it appears that its does not occur on its own).

Furthermore $*qa\dagger ud$ - $*qan'ud$ is a somewhat unstable base to build a proto-phoneme on. As pointed out in the list of cognates Amis $na-qalul$, Se $qəlul$, Ruk $arur$ all unambiguously point to a form $*qəlul$ 'flow'. In Amis $*qan'ud$ should reflect as $qa\dagger ur$, in Sediq as $qalud$ and Ruk as $?alud$. This means that $*qa\dagger ud$ - $*qan'ud$, the most widespread cognate is subject to doubling; a process which Ferrell (1982) describes as quite extensive in Paiwan. Therefore, this means that the Formosan forms and the non-Formosan forms could descend from different though phonologically similar doublets.

It appears that the reflexes of $*n'$ are very problematic and full of doubts for the few cognates which can be found on Formosa. We will therefore consider alternative solutions. The most obvious of these is to eliminate $*n'$ from the PAN phonemic inventory. If we do this then we are saying that $*n'$ arose at some period after the break-up of PAN. In section 3. a mechanism which would have caused this is discussed. Once n' was in the phonemic inventory it could easily be extended to forms with n (i.e. n becomes n' sporadically). Outside Formosa $*n$ and medial and final $*\dagger$ are always n and so with the cognates we have $*qa\dagger ud$, $*qa\dagger its$ (there is no evidence that outside Formosa $qa\dagger its$ was an analysable form) $*nala$ would be reflexed as $qanud$, $qanit$ and $nala$ which could easily become $qan'ud$, $qan'it$ and $n'ala$. There is other evidence for alternations between n and n' . In his discussion of the POC palatals Blust points out that the word $n'a[]$ or $na[]$ ($[]$ indicates that another syllable is usually added) which indicates 'mother' and various female relatives must be reconstructed with n' and n because of contradictory evidence. Minangkabau and Chamorro which usually reflex $*n'$ reflex this item with n , whereas Wogeo and some Admiralty languages reflex it with n' . Another case where Wogeo has a n' and a western language which usually reflexes $*n'$ as n' — in this case Malay — has n is tuna Ml, tun'a Wogeo 'eel'. Much the same as this example is the correspondence between Bare ρe $n'opu$ and Chamorro $nufuq$ (from $*nop(u)$ 'stonefish'). To solve the problems

with correspondence sets involving the POC root *natu 'child, offspring, ovary, egg' which has reflexes indicating a form n'atu, Blust proposes two roots *natu 'child' and *n'atu 'egg, ovary, baby bird'. With the view of *n' proposed here, the pressure to do this is lessened and while there may be other reasons to propose the two roots (and Blust does provide other good reasons), this is possibly another case of n' and n interacting. Finally there are two cases of n becoming n' when followed by i. The word for 'coconut' *niuɣ shows n' reflexes in many languages and the 3sg genitive *nia is reflexed as n'a in many languages. Therefore we have a situation where n' occurs when one would expect n, and where it appears that n' may correspond to two different phonemes in Formosa, which are reflexes as n outside Formosa in the cognates we have. These facts clearly argue that *n' arose after the break-up of PAN and in section 3. we will provide a mechanism for its becoming a phoneme. Therefore we do not reconstruct *n' as a PAN phoneme. This non-reconstruction accords with Dahl's and apparently Dempwolff's feeling that it did not occur in PAN.

Another of the phonemes which Dahl and Dempwolff suspected was not of PAN status is *k'. The only cognate Dahl cites for *k' on Taiwan is *k'aiŋ Pw tsaŋ 'to tether, to tie', Tg sahiŋ 'tar', Jv t'eŋ 'syrup', NgD saiŋ 'together'. However, as Tsuchida (1976:186 note 3) points out, Tg sahiŋ is from PAN t'ələŋ 'pine-tree, resin', so we can remove Tg. This leaves three forms with an extremely dubious semantic connection. It appears that we cannot accept these three forms as a basis for *k' cognate in Taiwan. There is however a *k' cognate in Taiwan (Zorc, personal communication). This is *pək'əq 'shattered, broken' Pw pətəq, Northern Tagbanwa pəsək, Jv pəcah, Ml pəcah, Fj voza, Sa pota. However, this is the only Taiwanese cognate known.

We should also consider some other facts about *k'. Outside a group of Western Indonesian languages (Javanese, Malay, Acehnese) and Malagasy, *k' always has the same reflex as *t'. In Malay, Javanese, Acehnese and the other Western Indonesian languages it has the reflex t' and in Malagasy it has ts (presumably from an earlier *t'). This is highly suggestive that what we are dealing with is not a separate phoneme but a sporadic retention of *t' in Western Indonesia. (It is generally accepted that the Malagasys came from Southern Borneo). Furthermore, we can provide a good reason why *t' should be sporadically retained as t'. Once *t' started to develop into s, a development which occurs throughout Indonesia and the Philippines, an unbalanced system of *d', *n' but no *t' would be left. To remedy this there are two solutions, either lose *d', as occurs in the Philippines, or sporadically retain *t' as t' in the non-basic vocabulary (an examination of the *k' cognates listed in Dempwolff 1938 shows that none of them occur in what could be called the basic vocabulary). As we have already mentioned, Malay, which we know has exerted a strong influence throughout this region, is in contact with the Austroasiatic languages which have a full palatal series in all positions. Therefore we do not accept *k' as a PAN proto-phoneme. To do so would be to set up a PAN proto-phoneme on the basis of a small number of non-basic vocabulary items which are largely confined to one area.

Therefore we have the following palatal phonemes *t', *d' and *g'. We should make some attempt to determine the phonetic nature of these phonemes. There is little doubt that *d' was d'; it has reflexes d, d' and s. *t' has reflexes t (Pw, Ruk, Th, Pl, Motu and Sa), s (most languages), ts (Amis), and dz (Pz). In the AN languages where nasal substitution occurs n' is the nasal which substitutes for *t' (if the language has n', otherwise n substitutes). (See section 3.). This argues that *t' was indeed a voiceless palatal stop.

The realisation of *g' is much harder to determine. It has reflexes d, ɣ, s, g, l, r, dz, n. As Dahl points out this tends to suggest a voiced palatal

affricate. However, this is unlikely because this would sound very like *d' which seems fairly clearly to have been d'. *g' could indeed have been a palatalised g or it could also have been a palatal fricative. A palatal fricative would be parallel to the apical s and velar γ. Phonological patterns are not however a particularly good reason to reconstruct a phoneme in a particular way. The reflexes are really too varied for the phonetics of *g to be reconstructed so we will continue to use the symbol *g'. This implies nothing about its actual phonetic realisation.

The final alteration which I would like to propose to the PAN phoneme inventory set up by Dempwolff is to consider the place of *g as a separate phoneme. The only cognates involving g that I am aware of in Formosa are gəmgəm ('fist, hit with fist' Dempwolff) git'git' ('cut into pieces' Pw gigit 'bite' Dempwolff) garuts ('comb' Pw, garut 'scratch' elsewhere) gatsəl ('itch' Pw, gatəl 'itch' elsewhere). However, garuts Paiwan is obviously derived from gutsguts 'scratch' and the infix ar 'having sound or quality of, involving use of'. Therefore, the word Dempwolff reconstructed as garut must be a frozen derived form from an earlier verb gut. Otherwise the Paiwan form could not be cognate. Therefore, we have three reduplicated forms *gəmgəm, *git'git', and a normal disyllabic root *gatsəl.

In the Philippines these forms all have doublets with k (Zorc, personal communication). This immediately raises questions as to their validity. Furthermore these roots all describe actions which are frequently described with onomatopoeic words, e.g. their meanings are 'hit with a fist, fist', 'cut into pieces, bite', 'scratch', and 'itch'.

There are a number of points about *g raised by Dahl (1976:86) which need to be considered. Firstly Dahl says that there are no instances of *g and *γ in the same wordbase. However, this is not correct, there is a root *tagəγaŋ 'ribs' reflected in Taiwan, the Philippines and Borneo (Blust, personal communication). This form has a k doublet (Zorc, personal communication) though. Dahl goes on to point out that γ occurs in only one reduplicated stem piγpiγ/miγmiγ 'spout' which has no Formosan cognates. On the other hand, Dahl points out that there are many instances of g in the same wordbase and not only in wordbases of the structure CVC².

Therefore it seems that *g has some sort of interaction with both *γ and *k. The question naturally arises as to whether there was a *g in PAN. The doublets with *k immediately suggest that *g is an innovation. However, there are no doublets in Paiwan, so one could equally validly maintain that the k doublets arose later in the Philippine languages. The occurrence of *g in words which are similar semantically, all having an onomatopoeic value and three out of four of which are CVC², and the non-occurrence of *γ in CVC² words makes one suspect that maybe *g and *γ should be combined. Against this we have the fact that *g and *γ are separately reflexes everywhere except in a few Philippine languages. Also we have *gatsəl, an ordinary disyllabic root.

From this evidence it appears that we must accept *g as a phoneme of limited occurrence with a phonaesthetic value. The status of *g in PAN would be analogous to the status of ŋ in present day English. *g was probably a fairly recent innovation as a phoneme in PAN, probably from *γ, it being the form of *γ used in CVC² words involving the idea of 'cut, rip, scratch' and then spreading from that start. It is also possible that it was the form of *k used in such words. Whatever its origins, it appears that *g was a PAN phoneme.

The last proposal concerning the PAN consonant inventory which we will consider is the one made by Dahl concerning *w and *j. Dahl demonstrates convincingly that *w and *u, and *i and *j were in complementary distribution in PAN. He does this by eliminating all sequences of *ij, *ji, *uw and *wu from Dempwolff's reconstruction. He shows that hiatus was the PAN thing found between combinations Vi, iV, Vu, uV. The languages which have j and w in these environments have inserted these glides. (There is usually no correspondence between words in the languages which insert the glides - a word with w in Jv does not usually have a cognate with w in Tg).

However, as Blust (1976) points out, the same complementary distribution applies in many present-day Austronesian languages but there are reasons to regard the non-syllabic i and u (j and w) as glides. Frequently failing to do so causes great difficulty in making statements about phonotactics and phonological processes in many languages. Furthermore, for example, Amblau has treated final j as a consonant in that it is lost like all other final consonants whereas final vowels are not lost. As Blust points out, adopting Dahl's proposal would obscure the dominant disyllabic character of the PAN lexicon. Therefore, I think until further evidence is uncovered the best solution is to regard *i and *j, *u and *w as being in complementary distribution but that there are phonotactic reasons for regarding *j and *w as glides.

Therefore the PAN consonant inventory we are proposing is:

p	t	ts	t'	k	q
b	d	ɖ	d'	g'	g
m	n				ŋ
	r				
	l, ɭ				
	s			y	h
w			j		

2. PAN GRAMMAR

2.1. PAN syntax and verbal morphology

There are three main syntactic types in the Austronesian languages. A focus system is found in the Philippines and Formosa; this is frequently called 'Philippine type' syntax. The two other types of syntax are 'Oceanic type' and 'Indonesian type' (Javanese, Malay). Philippine and Oceanic syntax are very different whereas Indonesian shares the features of both to a greater or lesser extent. The problem in reconstructing the clause structure of PAN has been which of these three syntactic types to reconstruct.

Foley (1976) argued that one should reconstruct PAN clause structure as being that of the Oceanic languages. Wolff (1973) appears to argue for the reconstruction of a pure Philippine style focus system. Pawley and Reid (1976) argue that PAN looked something like the Indonesian languages such as Toba Batak. Of these three approaches I consider Pawley and Reid's to be the one which explains the data best and which involves the most plausible changes.

Before discussing Pawley and Reid's proposal we will list the major properties of the Philippine and Oceanic clause structures within the framework of Role and Reference Grammar. Pawley and Reid present a description of POC clause structure which is generally accepted (Foley 1976). Therefore we will use this reconstruction, but rephrase it in terms of Role and Reference Grammar. The following are major features of POC clause structure.

- a) In transitive clauses Actor Verb Object is the unmarked word order.
- b) The verb consists of a stem and a clitic Actor concord pronoun and a tense/aspect marker which precede the verb stem. If the verb is transitive then, following the verb, there is a transitive suffix and an embedded pronoun which agrees with the Object.
- c) Actor and Object NPs have case marking whereas obliques occur as prepositional phrases.
- d) Actors are usually restricted to being agents, forces or experiencers.
- e) There are two transitive suffixes *i and *aki(ni). When *i occurs, this marks that the Object is a patient or product with an agentive verb, a stimulus or target with a psychological verb, and a location or goal with a verb of motion or posture. The transitive suffix *aki(ni) marks that the Object is an instrument with an agentive verb, a comitative with a motion or posture verb, and a causal or comitative with psychological verbs. It also marks that the Object is a benefactive.
- f) *i and *aki(ni) also occur as prepositions introducing obliques.
- g) Verb roots are divided into stative and active classes. Active verbs are divided into two classes; Action verbs (A verbs) and Process verbs (P verbs) (Arms 1974 - his analysis is of Fijian but it applies to POC). These verbs, when they occur as bare stems form intransitive clauses. To form transitive clauses, one can add the transitive suffixes *i and *aki(ni) to the stem. One can also form transitive clauses by adding the causative prefix *paka and *i or *aki. When an A verb is transitive the agent role, which occurs when it is intransitive, is the Actor. When a P verb is transitive the patient role, which occurs when it is intransitive, is the Object.

So far we have been using the term Object which is not a term in RRG without justification for so doing. The reason we are using the term 'Object' is that we need some term to cover the non-Actor nuclear role which occurs in transitive clauses in POC. Our use of the term implies nothing more than this. It is obvious that there are two types of Object, those marked by *i and those marked by *aki(ni). It appears that in POC all verb stems were monovalent and to make them bivalent, one had to add the transitive suffixes.

There are no such restrictions on the valency of verb stems in the focus languages. This is a correlate of the fact that clause structure is quite different in the focus languages from POC. Just how different it is can be seen from the following examples which show the four ways one can say in Paiwan 'the man hunts the pigs in the mountains with a spear'.

- | | | |
|----------|----|--|
| Actor | 1. | q/m/aʔup a tsautsau tau vavuy i gadu tua vuʔuq |
| Focus | | hunt man pigs mountains spear |
| | | 'the man hunts the pigs in the mountains with a spear' |
| Patient | 2. | qaʔup- n nua tsautsau a vavuy i gadu tua vuʔuq |
| Focus | | 'the pigs are hunted by the man in the mountains with a spear' |
| Locative | 3. | qa up-an nua tsautsau tua vavuy a gadu tua vuʔuq |
| Focus | | 'the mountains are where the man hunted the pigs with a spear' |

Instrumental 4. si-qaʔup nua tsautsau tua vavuy i gadu a vuʔuq
 Focus 'the spear is what the man hunted the pigs in the mountains
 with'

These four focus types are found in almost all focus languages and can be reconstructed for the Proto-Focus Language (PFL). The English glosses of these Paiwan clauses are not accurate reflections of the Paiwan. The English glosses imply that the Actor Focus (AF) construction is the basic construction. This is not the case with the Paiwan clauses. As can be seen from the clauses, the verb is equally marked in each case. This marking indicates that either the Actor is in focus, or the patient (PF), or the locative (LF), or the instrumental (IF). The NP in focus is indicated by a. None of these constructions can be said to be basic. They are all equally possible candidates for the basic structure. One can equally possibly derive the structure of the AF, PF and LF clauses taking the IF clause as basic, as one could derive the PF, LF and IF clauses taking the AF clause as basic, and so on. The NP in focus is the pragmatic peak (PrP) of the clause. The pragmatic peak is the centre of attention for the clause. In an IF clause, the hearer's attention is directed by the speaker to the instrument as being the most central participant role which is to be the pivot for the hearer's attention. In the focus languages any NP has an equal chance of being PrP with any other NP because of the four focus types. As a result of this system of clause structure, the properties associated with subjects (Keenan 1976) are split between the PrP and the Actor (Schachter 1977). The PrP controls referential properties such as relativisation, coreferential deletion and quantifier float, while the Actor controls role related properties such as reflexivisation (Schachter 1977).

These four clause structures, which occur in the focus languages, also occur in the perfective marked by an affix which is reconstructible for PFL. The perfective counterpart of the Paiwan IF clause given is:

s-in-i-qaʔup nua tsautsau tua vavuy i gadu tua vuʔuq
 'The man hunted the pigs in the mountains with a spear'.

However, though the focus system is very different from the POC clause structure, there are some common elements. The focus languages, like POC, have a class of stative verbs. In the focus languages and POC these verbs are marked by the prefix *ma-*. Some examples from Tagalog are *mabingi* 'become deaf', *mabasag* 'get broken', *mamatay* 'die'. Also found throughout Austronesia are the causative prefix **pa-~paka-* and the reflexive/reciprocal **may-*.

The Paiwan and Tagalog examples given here are representative of what can be reconstructed for the PFL and they show its major characteristics. These features still characterise almost all the focus languages and they are as follows.

- a) The basic word order is V NP*.
- b) One of the NPs is marked as PrP. Its case role is indicated by an affix on the verb. The other NPs normally have prepositioned role markers.
- c) Any nuclear role may become PrP.
- d) There are four focus types, Actor Focus, Patient Focus, Locative Focus, and Instrumental Focus.
- e) These four focuses may occur in the perfective when they are marked by **-in-*.
- f) There is a class of stative verbs marked by **ma-*.

Finally it should be mentioned that there is no strict division between nouns and verbs in most of the focus languages. Some morphemes such as personal names (e.g. John) are unlikely to occur as verbs. However, personal names can

be used as verbs meaning 'to call someone X'.

Wolff (1973) has a reconstruction of PFL which we will use. His reconstruction is as follows (using our terms):

	Non past	Past	Future	General	Imperative	Subjunctive
AF	-um-	-inum-				
PF	-ən-	-in-	r	-ən	-a	
LF	-an	-in- -an	r	-an	-i	-ay
IF	si-	s-in-i			-an	

r = reduplication of first CV of base.

We have made one alteration to Wolff's forms. Wolff has *i* not *si* for the IF prefix. Taiwanese evidence clearly shows that it was **si*. This should be reflected as *hi* in the Philippines, however apart from Tausug *hipag-* and Waray *mahi-* (see Zorc 1977:134), it is always reflexed as *i*. However the Tausug and Waray evidence clearly require us to reconstruct **hi* for PPh, as a reduction **hi* → **i* is one of the most natural sporadic sound changes in the world.

The absence of *-*ən* in the perfective PF calls for some comment. This absence of *-*ən* is usually explained as being due to 'zeroing-out' whereby the presence of *-*in-* caused the phonologically similar suffix *-*ən* to disappear. However, this explanation is totally inadequate, for if *-*in-* causes *-*ən* to disappear, it should also cause *-*an* the LF marker to disappear. There is a much better alternative explanation. The PF construction is an ergative construction. The connection between ergativity and perfectivity is well known. In Tongan ergative clauses are perfective and accusative clauses are non-perfective (Foley 1976). In some Caucasian languages perfective clauses are ergative and non-perfective clauses are accusative (Dixon 1979). Therefore it is reasonable to suggest that if a perfective clause carries no focus marker for which role is PrP then the unmarked reading of a clause with *-*in-* only would be that the patient was PrP (i.e. that it was a PF or ergative clause).

So far we have been presenting a picture which views the focus affixes as simply conveying the information that role *x* is the PrP of the clause. This is a rather simplistic view of the function of the focus affixes. Wolff (1973) shows that in PFL, the LF marker was used for patients of accidental and involuntary actions. The examples below show this:

SL *tingali hibati?an ?a ?atu? pagharampang*
lest hear-LF CM= our conversation
'Maybe someone will overhear our conversation'

Ts *oana i si nana ahta talui*
no longer TM by her particle at all happen to be heard
'She never heard it again'

The LF marker was also used to refer to recipients of non-volitional verbs such as forget. The LF marker was used to refer to the thing with reference to which a mutual action takes place, whereas the people engaged in the action were referred to by PF.

SL *pag-?a-?away-an ta ?a nahinabu?*
fut-quarrel-LF CM= what happened
'We will quarrel about what happened'

SL ʔa-ʔaway-un ku siya
 fut-fight-PF I he
 'I will fight with him'

Wolff states that in many languages the LF marker has extended its range at the expense of the PF marker.

In PFL the instrumental focus marker also covered beneficiaries. It also covered the recipients of verbs of motion away from the Actor (i.e. giving, throwing). The IF marker is also used with verbs such as say, tell, etc. to refer the thing asked, etc. Working on the principle that affixes lose semantic specificity over time, it is probable that these affixes had even more concrete meanings in PFL. Foley (1976) in his analysis of Tagalog shows that in present-day Tagalog one cannot treat the affixes as having only a pure case-marking function. The focus affixes have a discourse function. In most languages, old information occurs clause initially for linkage purposes, but in Focus languages the verb which is new information occurs initially.

The focus affixes, however, refer to the PrP which is in all focus languages usually old information. From the semantics of the affix and the case frame of the verb one is able to tell which NP will be the PrP and so one knows which pieces of information from the previous clause(s) is the centre of this new clause.

Finally in our discussion of the focus affixes we should note that it appears they could all be used as nominalisers in PFL. This fact will prove quite important in our discussion of PAN syntax. Dardjowidjojo (1978), in his discussion of nominal derivation, states that Kə, pəN, pər, -an and combinations of the first three and -an occur as nominalisers. Now -an is obviously a reflex of the LF marker *-an. It forms nouns with the meaning of (i) what one (verb), (ii) things which are (adjective), (iii) place where one (verb), (iv) collection of X. Examples are (i) anjur-an 'suggestion', (ii) manis-an 'sweets', (iii) parkir-an 'parking space', (iv) laut-an 'ocean'. The third use also occurs in Paiwan (Ferrell 1982) and Tagalog (Schachter and Otnes 1972) languages where -an still functions as the LF marker. PəN occurs as focus affix in the Philippines but is not reconstructable for PFL (see section 3.). Kə, on the other hand, is found throughout the focus languages, both alone and in combination with -an. In Indonesian kə- is a frozen form occurring in only three items. In Paiwan its uses are varied (e.g. ka-tiaw 'yesterday', si-ka-taqəd 'reason for sleeping'). In Tagalog, the time words ka-hapon 'yesterday' and ka-gabi 'tomorrow' take it as they do in Paiwan. Tagalog also has forms such as ka-kwarto 'room mate' (kwarto 'room'), karelihyon 'co-religionist' (relihyon 'religion'), katabi 'someone/thing beside' (tabi 'side'). In Palauan klə, from *k-in-a with the perfective infix, forms abstract nouns from stative verbs (Joseph 1975). It also functions as a verbal affix in Paiwan and Tagalog. In Paiwan it functions as an inchoativiser for some stems (Ferrell 1982) and in Tagalog it occurs alone or in combination as Maka-. Alone it means involuntarily doing something; maka also indicates this but it also indicates the ability to do the action of the verb. The meanings attached to this prefix are rather diverse but basically it seems to have two meanings. These are 'state' and 'unmonitored, uncontrolled change of state'.

In Paiwan, Tagalog and Indonesian we also find *ka- ... -an as a nominaliser. This form does seem to have a reconstructible semantic value. In Paiwan Ferrell says that its meaning is 'principal or main manifestation of'. In Tagalog Schachter and Otnes define it to mean 'the class or qualities or group of people/things instanced by what the base designates'. In Indonesian Dardjowidjojo

gives its meanings as (i) product of (ii) state or quality of being (iii) location (iv) collection (v) abstraction. From these descriptions one can see that there is a common idea of 'group or individual which manifests the thing instanced by base'.

In Tagalog and Paiwan *ən can also be a nominaliser. Examples are Pw kən-ən 'food' (kan 'eat'), Tg awit-in 'song' (awit 'sing'). It is used as a nominaliser in Palauan. In Paiwan the AF marker -m- and the IF marker si- can be used as nominalisers, e.g. q/m/uʔuts 'wrestler': quʔuts 'wrestle', si-qunu 'knife': qunu 'cut'. The AF and IF markers can also be used as nominalisers in Tagalog. (e.g. ang pumasok 'the one who entered', ang ibibili niya 'the thing he is buying'). The perfective infix *-in- also occurs as a nominaliser throughout the Philippines.

Therefore it appears that in PFL the focus affixes were bound morphemes which had a definite semantic value which conveyed information about the NPs associated with the verb. They could also be used to form nominals from verb stems. The semantics of these nominals was probably determined by the semantics of the focus affixes.

Having reconstructed these affixes and fitted them into the focus system, it is of great significance for any reconstruction of PAN syntax that we find the same affixes functioning as nominalisers in languages with an Oceanic syntax. Pawley and Reid (1976) point out that *-an, *si- and *-in- are productive and reflected in many Oceanic languages as nominalising affixes and *-ən has traces in the Oceanic languages. There are also traces of *m- in Oceanic (there is replacement of initial p or b by m sporadically - in many focus languages when a stem with initial p or b occurs in AF, the p or b becomes m). The semantics of the Oceanic affixes are relatable to the semantics of the affixes in the focus languages.

The next major fact which must be accounted for in any reconstruction of PAN syntax is the presence of reflexes of the Oceanic transitive suffixes *i and *aki(ni) in Indonesian. In Indonesia these suffixes indicate forms *i and *akən. Pawley and Reid (1976:60, 1976:70 note 15) provide strong arguments that *akən is the original form of the suffix. These suffixes have uses which match those of Oceanic languages. However, languages such as Toba Batak and Wolio which have these suffixes make a three-way distinction against a two-way one in Oceanic. In Toba Batak those roles which are marked by Patient Focus in the focus languages occur with a \emptyset marked verb (no suffix), those which are marked by LF, are marked by i in TB and those which are marked by IF are marked by -hon. In Oceanic *Ci covers both TB i and \emptyset .

We may assume that the TB represents the original situation. It is more likely that Oceanic would have extended the range of *i than that TB and to an extent Wolio would have restricted it. The theory which best explains the presence of the focus affixes in the Oceanic languages and the presence of the Oceanic suffixes *i and *akən in Indonesia is Pawley and Reid's (1976). They propose that PAN had clause structures with both Oceanic and focus features, a situation which occurs in languages such as Toba Batak. The reconstruction, which we propose, differs in details from that of Pawley and Reid.

We propose the following clause structures:

1. Stative clauses.

ma-Verb NP*

2. Active clauses.

a) AF clauses

V-um-erb NP_{Pat} ACTOR NP_{Loc} NP_{Inst}
 V-um-erb-i NP_{Loc} ACTOR NP_{Inst} (NP_{Pat}?)
 V-um-erb-akən NP_{Inst} ACTOR NP_{Loc} (NP_{Pat}?)

The perfective has -inum-.

b) PF clauses

Verb-ən NP_{Pat} ACTOR NP_{Loc} NP_{Inst} PP* (non-perfective PF)
 V-in-erb NP_{Pat} ACTOR NP_{Loc} NP_{Inst} PP* (perfective PF)
 Verb-an NP_{Loc} ACTOR NP_{Pat} NP_{Inst} PP* (non-perfective LF)
 V-in-erb-an NP_{Loc} ACTOR NP_{Pat} NP_{Inst} PP* (perfective LF)
 si-Verb NP_{Inst} ACTOR NP_{Pat} NP_{Loc} PP* (non-perfective IF)
 s-in-i-Verb NP_{Inst} ACTOR NP_{Pat} NP_{Loc} PP* (perfective IF)

The development of the Oceanic type syntax would follow the lines suggested by Pawley and Reid. PAN would presumably have had a verb-initial clause structure with free order among the NPs. As has been already discussed the function of focus affixes is to provide old information for discourse cohesion purposes when each clause starts with new information. The basic innovation in Oceanic was the movement of the Actor to the first position in this clause. It seems that the best explanation for this is either that PAN had an alternative construction for sentence-level topics of the form Topic Verb (NP*) with a pronominal trace of the Topic in the clause as occurs in many languages throughout the world, or that some high-order subgroup (POC?) innovated this. It seems most likely that it existed in PAN as it appears that all languages have sentence-level topicalisation procedures and one leaving a pronominal copy in the clause appears to be the commonest cross-linguistically. One of the main functions of topicalised constructions is to introduce new themes in discourse. From this start it is relatively simple for it to spread to introducing themes in general. Once this NP initial construction became an unmarked construction there would be little reason to maintain the focus system, because there would be an alternative system with old information clause initially. Parallel to, or following, this would have a restriction in the roles which could occur in this clause-initial position until only Actors could occur there. This is an unsurprising development considering that Actors tend to act as themes for discourse in the majority of the world's languages. Once these two things had happened we would have clauses with the structures:

Actor Verb \emptyset Patient/Goal X
 Actor Verb i Locative X
 Actor Verb-akən Instrumental X.

To obtain the Oceanic clause structure we only need to collapse the \emptyset and *i categories. This can simply be regarded as example of the spread of locative markers into the range of patients and goals. The non-Actor focus constructions have been lost in Oceanic. The focus affixes have been retained in their nominalising function. The reasons for the loss of the non-AF constructions are not entirely clear. As we have pointed out the development of noun-initial word order would make the focus affixes obsolete to a large extent. The AF marker

m- would have no function because the Actor would always occur initially, and so it is lost almost without trace in Oceanic. When the word order in Oceanic became Actor initial we would have two conflicting systems of discourse linkage with the non-AF clauses. We would have one system which keyed on Actors and another on the focus affixes. Apparently in Oceanic the system of linkage keying on Actors won out and the system of linkage on focus affixes was lost and so the non-AF clause types were lost.

There is one further thing which we must explain in our derivation of POC clause structure. The POC Actor concord pronouns correspond to the PAN PrP pronouns and the Object concord pronouns correspond to the PAN genitive pronouns which covered genitives and non-focussed Actors. As the Actor was PrP in the AF constructions, the fact that the Actor concord pronouns are the old PrP pronouns is to be expected. However, as we shall see, the genitive pronouns did not cover patients in PAN. In some focus languages, though, such as Tagalog, non-focussed actors, patients and genitives form a category (Schachter and Otnes 1972). If this was the case in POC then the use of these genitive pronouns as object concord pronouns in POC is to be expected because the patients would have been non-focussed.

This is how we would derive the Oceanic active transitive clause system described earlier. The intransitive clauses would develop simply from the PAN monovalent clauses by the topicalisation procedure. In this there could be no restriction of initial position to Actors only because some verbs (the process verbs) by their semantics require patients.

This course of development follows the course proposed by Givón in his article "Topic, pronoun and grammatical agreement" (1975), where he says that "when a language reanalyses the topic constituent as the normal subject or object of the neutral, non-topicalized pattern, it perforce also has reanalysed subject topic agreement as subject agreement and object topic agreement as object agreement". We are saying that this is what happened in the development from PAN to POC.

A development similar to this appears to be going on in literary and Manila Tagalog today. In her discussion of Tagalog, Naylor (1975) says that verbs with the IF marker are not used. She also states that Benefactive focus forms were much rarer than she expected and that she could find only four instances of Locative focus constructions in all her texts. The rarity of these constructions in written texts may be related to the extensive use of topicalised 'ay' constructions which have initial NPs, in written texts. The use of ay constructions has not penetrated into ordinary colloquial speech, and in Manila Tagalog the disuse of the focus constructions mentioned above occurs and so there is apparently no necessary connection between loss of specific focus types and the rise of topicalised constructions. The factor which argues that there was such a connection in the development of Oceanic syntax are the obligatory pronominal copies which are as Givón argues concomitants of topicalised constructions.

The development of all the focus languages is relatively simple. The focus languages have simply made AF constructions have the same options as the other types of focus. This gets us to the situation we find in most of the focus languages.

Taiwanese Paradigms (not necessarily complete).

Kanakanabu (Tsuchida 1976)

		Enclitic		
		Nominative	Genitive	Oblique
sg 1	i:ku	-(m)aku	-(m)aku	-ʔikua
	2 i:kasu	-kasu	-(mu)su	-kasua
pl li	i:kita	-kita	-(mi)ta	-kitana
	le i:kimi	-kimi	-mia	-kimia
2	i:kamu	-kamu	-mu	-kamua

(mV) occurs after vowels

Saaroa (Tsuchida 1976)

		Enclitic		Oblique
		Nominative	Genitive	
sg 1	iʔaku	-aku	-ku	na + independent forms
	2 iʔaa	-u	-u	
pl li	iʔaʔamu	-amu	-ʔamu	
	le iʔaʔa	-ita	-ta	
2	iʔamu	-mu	-mu	

Tsou (Tsuchida 1976)

		Enclitic	
		Nominative + Genitive	
sg 1	aʔo	-ʔoʷʔu	
	2 su:	-koʷsu	
pl li	aʔmi	-mza	
	le aʔto	-to	
2	mu:	-mu	

Thao (Li 1978)

		Nominative	Non-Focussed NP	Genitive (Possessor only)
sg 1	ya:kuʔ	ya:kin	ya:kin	na:k
	2 ʔi:huʔ	ʔu:hun	ʔu:hun	mi:huʔ
	3 ʔi:θuʔ	ʔi:θuʔ	ʔi:θuʔ	ʔi:θuʔ
pl li	ʔi:taʔ	ʔu:tan	ʔu:tan	mi:taʔ
	le ya:min	ʔa:min	ʔa:min	na:m
	2 mani:yun	ʔi:human	ʔi:human	mani:yun
3	θayθuy	θayθuy	θayθuy	θayθuy

Kavalan (Li 1978)

	Nominative	Genitive
sg 1	-iku?	-ku?
2	-isu?	-su?
3	∅	-na?
pl li	-ita?	-ta?
le	-imi?	-nyaq
2	-imu?	-numi?
3	∅	na qanyau?

Pazeh (Li 1978)

	Nominative	Genitive	Oblique/Topic
sg 1	-aku?	naki?	yaku?
2	-siw	nisiw	?isiw
3	-misiw	nimisiw	?imisiw
pl li	-ta?	nita?	?ita?
le	-ami?	niami?	yami?
2	-mu?	nimu?	?imu?
3	-amisiw	na:misiw	ya:misiw

Saisyat (Li 1978)

	Nominative	Genitive	Oblique
sg 1	yako ^v yao	ma?an	yakin
2	ʃo?o ^v ʃo?on	ni-ʃo?	?i-ʃo?on
3	sia	ni sia	hi sia
pl li	?ita?	mita?	?ini-mita?
le	yami?	ni-a?om	?ini-a?om
2	moyo	ni-mon	?ini-mon
3	la-sia	na-sia	hi la-sia

Nataoran-Amis (Teresa M. Chen, personal communication)

	Topic	Nominative	Genitive	Locative	Benefactive
sg 1	ako	kako	nomako (naku)	itakoan	
2	iso	kisu	nomiso (niso)	itisoan	saka
3	cira	cira	nira	iciraan	+
pl li	ami	kami	noniam	itamian	Locative
le	ita	kita	nomita (nita)	ititaan	
2	amo	kamo	namo	itamoan	
3	oqni	koqni	noqni	itoqnian	

Paiwan (Ferrell 1982)

	Nominative	Genitive	Category 3
sg 1	-akən, ti-akən	ku-, ni-akən	t'anu-akən
2	-sun, ti-sun	-sun-, ni-sun	t'anu-sun
3	ti-mad'u	ni-mad'u	t'ai-mad'u
pl li	-it'ən, ti-t'ən	t'a, ni-t'ən	t'anu-it'ən
le	-amən, ti-amən	nia-, ni-amən	t'anu-amən
2	-mun, ti-mun	nu-, ni-mun	t'anu-mun
3	tia-mad'u	nia-mad'u	t'aia-mad'u

Rukai (Li 1975)

	Independent		Dependent	
	Subject	Object	Subject Marker	Personal Possessive
sg 1	ko-n-ako	n-akw-a	-(a)ko, -n-ako, naw-	-li
2	ko-so	mo-sw-a	-so	-so
3 def	(ko-a-ni)	(i-ny-a)	∅	-i-ni
3 indef	(ko-a-da)	(i-da)	∅	-i-da
pl li	ko-ta	mi-ta	-ta	-ta, -i-ta, -mi-ta
le	ko-nai	nay-a	-nai	-nai
2	ko-nomi	nomy-a	-nomi	-nomi
3 def	(ko-l-i-ni)	(l-i-ny-a)	∅	-l-i-ni
3 indef	(ko-l-i-ḍa)	(l-i-ḍa)	∅	-l-i-ḍa

(mon 2pl Li 1975:257 archaic form).

Atayal (Egerod 1966)

	Nominative	Genitive	Locative	i forms
sg 1	saku? ku?	maku? mu	knan	kuziŋkʷkun
2	su?	su?	(i)sunan	isu?
3	hia?	nia?	hian	hia?
pl li	ta?	ta?	?tan itan	ita?
le	sami	mian	sminan	sami
2	simu	mamu	smunan	simu
3	hga?	nha?	hgan	hga?

Sediq (Asai) [This paradigm is not accurately or fully recorded]

	Nominative	Genitive	Locative
sg 1	jakuzku	nakuʷmu	kenan
2	isuʷsu	nisuʷsu	sunan
3	hiḍa	niḍa	
pl li	yami	mian	
le	ita	nitaʷta?	
2	yamu	namu	
3	dahiḍa	nuhiḍa	

Bunun (Jeng 1977)

	Nominative	Oblique + Contrastive	Accusative enclitic
sg 1	sak ^u ?a ^o ak	o ^o aku	-ku
2	?asu ^u ? ^u su?	su ^u ?u	-su
pl li	?ata?	?ita?	-ta?
le	sam ^u ?a ^o am	o ^o ami?	
2	?amu?	mu ^u ?u	-mu
sg 3	sia	?ista	-ta?
	siata?	?istun	
	?aipa? (far)		
	?aipi? (near)		
	?aipu? (far)		
	?aipun		
pl 3	nai	?inta	
	naip	?intun	
	naita?		
	?iŋka?		
	?iŋkun		

Genitive

	pre-nominal	post-nominal
sg 1	inak	nak
2	isu ^u ?u	su ^u ?u ^u su
3	isiata?	?itu?
pl li	imita?	mita?
le	inam	nam
2	imu ^u ?u	mu ^u ?u ^u mu
3	inaita?	naita?

2.2. PAN pronouns

We will now go on to consider the pronominal system which can be reconstructed for PAN. As the system we have reconstructed for syntax is basically that of a focus language, we would expect that the pronouns would have categories similar to those found in the focus languages. The focus languages have essentially three categories for pronouns. The categories are widely called nominative, genitive and oblique. The nominative form is used when the pronoun is in focus. The genitive covers non-focussed actors as well as genitives, while the oblique covers all other pronoun functions including absolute genitives 'mine, yours' etc. The PAN pronouns which we reconstruct for these categories are not necessarily reflected in these categories nor are the categories themselves necessarily reflected. We will reconstruct four sets of pronouns; one for the nominatives, two for the genitives, one a bound set and the other free, and we will reconstruct a fourth set which may or may not have covered all the range of the third category.

The fullest discussion of the PAN pronouns to date is Blust 1977. He proposes the following paradigm:

	nominative	genitive
sg 1	i-aku	i-ku/ni-ku
2	i-su	i-su/ni-su
3	t'i-ia	i-a/ni-a
pl 1i	i-kita	i-ta/ni-ta
le	i-kami	i-mi/ni-mi
2	i-kamu	i-mu/ni-mu
3	t'i-ida	i-da/ni-da

We will examine the forms of the genitive pronouns first as these appear to have undergone the least alternations since PAN. Also evidence from the genitive paradigm will be important in determining the nominative paradigm precisely because the genitives have undergone less change. Blust's genitives correspond to our free form genitives. Blust's genitives in *ni- are reflected in many languages such as Pazehe and Thao. Blust's reasons for reconstructing the *i genitives are that i occurs as genitive article, as well as ni, in Raluana, Bugotu (Oceanic languages), various Batak languages (Sumatra) and Tontemboan (Sulawesi). He finds support for it in alternations between nX genitives following vowels and X genitives after consonants in Javanese, Proto-Ambon, Buginese and Toba Batak.

However, the only form Blust cites which must descend from an *i genitive is Javanese e (from *ia) postconsonantal (ne postvocalic) 3rd sg genitive. The other alternations kuŋku (1st gen) and taŋta (1pl incl) provide no evidence for an i genitive. Furthermore, as we will see in our discussion of the nominative forms, it is doubtful whether PAN had third person pronouns. In any case the existence of postvocalic nV and postconsonantal V variants of suffixed genitive pronouns in one language is not a firm basis for positing that PAN had post-vocalic ni and postconsonantal i genitives. The Javanese e form could have arisen from a reduction of Cne to Ce. However, Reid (1979) states that there is evidence from Bontok to support Blust's -i genitives. Therefore for the moment we will regard the status of Blust's i genitives as uncertain.

On the other hand the genitives reconstructed by Dahl (1976) and Dyen (1974) are reflected in many languages. The forms reconstructed are:

sg 1	ku
2	su
3	n'a
pl 1i	ta
le	mi
2	mu
3	da

Forms such as these occur in Tsou, Saaroa, Javanese and Magori and many other languages. These languages do not necessarily have all these pronouns but on the basis of many languages the above pronouns can be reconstructed.

Blust makes a number of points about this paradigm. The first is that *n'a is better reconstructed as *nia, a form which is segmentable into *ni the well known genitive marker and *a an element which occurs in the third person nominative pronouns. The second is that outside Taiwan the 2nd pl clitic mu is frequently reflected as a 2nd sg. As Blust points out, this is probably due to a politeness shift, a well known cross-linguistic phenomenon. The third point which we will note is that there is no evidence in Taiwan for a 3rd pl genitive *-da and therefore we will not reconstruct it for PAN.

A number of Taiwanese languages provide evidence for another genitive. Saisyat maʔan, Atayal makuʔ and Kananabu (m)aku (m after vowels) provide evidence for a 1st sg genitive in m. However, these three forms are not relatable. While Atayal and Kananabu point to *maku, Saisyat cannot. In Kananabu the 2nd sg and 1pl incl, (mu)su and (mi)ta have m alternatives following vowels. Saisyat and Bunun also have mitaʔ 1pl incl gen and Thao has mi:huʔ 2sg gen. These forms all argue for some sort of m genitive. However, its form and functions are now irretrievably lost. It may have been the original pre-PAN genitive now left only in traces.

The form and function of the ni pronouns is on the other hand recoverable. While we agree with Blust that there were ni pronouns, we do not agree with him as to their form or functions. Pazeh nakiʔ, Thao na:k, Bunun (i)nak, (1sg gen) Pazeh niamiʔ, Thao naim, Bunan (i)nam, Saisyat ni-aʔam (1pl excl) all point towards forms *ni-aku (1st gen) and *ni-ami (1pl excl gen) consisting of *ni the genitive marker and a nominative root. As the 1pl incl nominative root corresponding to *aku and *ami is *ita we will get *nita and we cannot say from this form whether it is *ni-ita or *ni-ta. However, *ni-ami and *ni-aku point towards *ni-ita unequivocally. As the second person nominative roots are *su and *mu we will get *ni-su and *ni-mu anyway. Therefore we reconstruct the two genitive pronoun paradigms as:

sg 1	ku	ni-aku
	2	su
	3	nia
pl 1i	ta	ni-ita
	le	mi
	2	mu

The question naturally arises as to what the differences between the two sets was. The difference, as we have indicated, is that the ni set are free-form pronouns and the other set are bound. Dyen (1974) reconstructed the first set as clitics and it can be seen that they are bound forms in the Formosan languages where they occur. On the other hand, the ni forms when they occur in Pazeh, Thao and Saisyat are free forms. The Bunun i genitives are not related to Blust's proposed *i genitives. They differ in form and function. Jeng (1977) says that they occur after the attributive particle tuʔ and preminally with the consonant forms occurring postnominally.

We will now go on to discuss the pronouns for the third category. These pronouns are reflected in many languages such as Paiwan, Pazeh, Ivatan, Tagalog, Okolod Murut, Cebuano and Malagasy (see Blust 1977). The forms that can be reconstructed are:

sg 1	akən
	2
pl 1i	itən
	le
	2

These forms are composed of a corresponding nominative root and the PF suffix *-ən (Blust 1977). The combinations are:

sg 1	aku + ən
	2
pl 1i	ita + ən
	le
	2

The difference in results of adding *ən* to the first and second person forms is probably best explained by a stress difference. The first person forms are disyllables and the second person forms are monosyllables. If the disyllables were initially stressed then the difference is explicable.

These forms have many functions in the various languages where they occur. In Tagalog they are the oblique forms; in Paiwan the base; in Okolod Murut the goal focus forms. However, in all languages where they occur, either the base alone or some affixed form of this base (whatever other bases occur) covers the meaning of the absolute genitives 'mine, ours', etc. Therefore this function must as Blust points out be attributed to these forms in PAN. This function follows naturally from the semantics of a combination of the root and the PF marker **ən*. Cross-linguistically the connection between 'It is to me' and 'It is mine' is well known. This pronoun form presumably also is an oblique form at least for patients and goals. Therefore the use of this form as an absolute genitive is unsurprising. We do not know if this form covered all the range of the third category in PAN. This combination of a pronoun root and a focus affix may seem odd but in Atayal and Sediq there are forms of the pronoun which consist of a root and the LF marker *-an* (Egerod 1966). There are similar forms in Amis. These forms are the oblique forms. In Thao and Saisyat these **ən* forms function as obliques for the 1st sg and in Thao for the 2nd sg as well. Furthermore, as we have already seen, though there are verb and noun slots, the fillers are not so categorised normally and it appears that in PAN 'verbal' and 'nominal' affixes (the PF marker *-ən* and as we shall see the noun focus particle *i*) could be affixed to the pronouns. In Paiwan these forms have completely replaced the original nominatives and are now bases for a restructured paradigm. Reflexes of **akən* as the first person nominative are found in many languages. In Tagalog *amin* and *atin* (from **amən* and **itən*) are the non-nominative bases.

It appears that there is a tendency for these forms to extend their range. One possible explanation might be the use of these forms in cleft-like constructions such as 'it was me who did it' where the oblique form is the focussed form of the lower clause. This usage could explain how these oblique forms spread.

We will now go on to consider the forms of the nominative pronouns which Blust reconstructed as:

sg 1	i-aku	
2	i-su	i-kasu (polite)
3	t'i-ia	
pl 1i	i-kita	
le	i-kami	
2	i-kamu	
3	t'i-ida	

We have left the discussion of these forms to last because they show the most variation and are therefore the most difficult to reconstruct. Also paradigmatic evidence from the three paradigms we have already reconstructed will be important. We will consider the forms of the pronouns separately. The form of the first singular is not subject to much dispute. The form reconstructed by Brandstetter and Dempwolff was **aku*, reflexes of which are found in many languages such as Tsou, Javanese, Malay, Tagalog, Fijian and Samoan. A form **i-aku* with the prefixed common noun focus marker **i* is found in many languages such as Thao, Saisyat, Bareʔe, Bugis and Jandena (Blust 1977). Kavalan and Kananabu *ikuʔ* and *i:ku* are presumably further developments from this. Atayal *sakuʔ* corresponds to Sediq *yaku* and Atayal *sami 1pl excl* and *simu 2pl*

correspond to Sediq yami and yamu. As Sediq and Atayal form a subgroup (see p.89) it appears that one should reconstruct Proto-Atayalic *yaku. The same explanation may also hold for the Bunun forms sak^v?a^oak. In Bunun $y \rightarrow \check{o}$. These forms could also represent *t'i-aku with the proper noun focus marker prefixed. This form should give siak in Bunun but a change siak \rightarrow sak is not particularly unlikely. In Ilokano a form siak, siaku-n occurs which could also derive from *t'i-aku. However, it is also possible that it derives from Proto-Cordillerean *si-yakən (Reid, personal communication). There does not appear to be enough evidence to posit the existence of a PAN *t'i-aku, as sak in Bunun could possibly arise from *yak as it does in Atayal.

The second person is the most problematic one in AN pronominal reconstruction. The reflexes indicate three different roots *ka, *i-su and *mu which combine in many different ways and appear to swap over in number in the different languages. We will consider the singular forms first. *ka is found in Tsou and many Philippine languages. *i-su is found in many Formosan and Philippine languages and in Malagasy. This appears to consist of the focus marker *i and a form *su which occurs in Tsou, Bunun and Atayal. A combination of the two *kasu is reflected in many non-Formosan languages. A form *ikasu is reflected in Kananabu, Tagalog, Fijian and other languages. This means we have forms *ka, *su, *i-su, *kasu, *i-kasu for the 2nd sg. Blust proposes that *(i)-kasu was the polite pronoun in PAN with *(i)-su being the familiar one. There is, however, no language which reflects *(i)-kasu as a polite pronoun and this explanation does not tell us what the functions of *ka and *(i)-su were. In Tsou ko (*ka) and su (*su) occur in morphophonologically conditioned environments; su occurs after a focus particle ending in o or oh and ko elsewhere. However, it would be exceedingly dangerous to say that this reflects the PAN usage. Finally it must be mentioned that Ivatan has a singular pronoun ?imo which corresponds to a plural elsewhere. It is generally assumed that the Ivatan use arose as a result of a politeness factor. Tagalog reflects both *i-kasu and *ka without apparently there being any particular difference in politeness.

I think that one must also consider the plural forms *kita, *kami and *kamu. Dahl (1976) has pointed out that the Formosan languages appear to reflect forms without the k and that the k may reflect some sort of old pronominal prefix. If we examine the Formosan languages we find, corresponding to *kita, Bunun ?ata? and ?ita?, Kavalan ita?, Thao ?i:ta?, Saisyat ?ita?, Paiwan it'ən, Atayal ita?, Saaroa i^oata, Rukai ko-ta and Tsou a^oto. These forms point towards *ita but Tsou, Rukai and Bunun also indicate *ata (in Tsou initial k is lost so the Tsou form could reflect *kata). If we examine the reflexes of the 1pl excl in Formosa we get Bunun sam^v?a^oam, Kavalan imi?, Thao ya:min, Saisyat yami?, Paiwan amən, Atayal sami, Pazeh yami?. These forms all point to *ami. Finally if we examine the forms for the 2nd pl we find Bunun ?amu?, Archaic Rukai mon, Paiwan mun, Atayal simu, Pazeh ?imu?, Saaroa i^oamu, Tsou mu:, Kavalan imu?, Saisyat muyu. These forms indicate a form *mu (the i which is found here and on the Formosan forms for kita and kami is of course the noun focus marker). It is not only the Formosan languages which lack the k. Ivatan (Reid 1971) has imo (2nd sg - corresponding to 2nd pl in Formosa), Subanon has yami, yamu and ita.

The parallelism between *su: *kasu and *mu: *kamu is too obvious to miss. The similar parallelism *ita: *kita and *ami: *kami is also obviously the same phenomenon. It appears therefore that an element *ka which is found as a 2nd sg pronoun could be prefixed to the other 2nd sg *su, the 2nd pl *mu, the 1pl incl *ita and 1pl excl *ami. The common factor for the prefixing of *ka seems to be

the 'second person', with *kami being due to paradigmatic pressure from *kita and *kamu.

There is, however, further data which must be brought into consideration when dealing with the question of *ka and the second person. Saisyat moyo, Itbayatan imi:yu, Sambal muyu, Casiguran Dumagat moy and Agta moy (Reid 1971) all point toward a 2nd pl pronoun *muyu. Ivatan has a form ?inio?, Binuhid has sinyu and Tagalog has a benefactive form ?inyo which point to a form *inyu. Now reflexes of a form *kamuyu are found in AN languages such as Agta. Reid (1974) has reconstructed a form *kamuyu for the 2nd pl in Proto-Cordilleran (a Philippine subgroup). A form kayu occurs in a number of Cordilleran languages as well as in the neighbouring Tagalog and Kapampangan. Harmon (1974) in her reconstruction of Proto-Manobo (another Philippine subgroup) has an emphatic form *kiyu which is presumably related to *kayu. Finally in Tagabili the 2nd pl pronoun is yuh (Reid 1971). All of these facts suggest that there was a 2nd pl pronoun or pronoun suffix *yu.

All of these facts combine to make determination of the forms of the second person pronouns in PAN rather problematic. For the 2nd sg we have *ka, *su and *kasu; for the 2nd pl we have *mu, *yu, *muyu, *kamu, *kayu, and *kamuyu. If we accept that the *ka forms of these pronouns arose as a result of prefixing *ka 2nd sg to stems, then we must regard the non *ka forms as older. Support for this is provided by the fact that in the paradigms we have already reconstructed there is no trace of *ka.

This leaves us with *ka and *su for the 2nd sg and *mu, ?*yu and *muyu for the 2nd pl. This *yu form is rather problematic. Outside Formosa *s is always reflected as either h or \emptyset and so *i-su gives i(h)u. Now Tagalog has a form iyo (*iu) which is the locative form of the 2nd sg and Malagasy has a form iu which was a 2nd sg pronoun. The question arises as to whether this iu from *i-su is the same as the *yu we have reconstructed as occurring in the 2nd pl. Blust (1977) argues that they are, but as he points out we cannot explain the Saisyat form moyo 2nd pl this way. In Saisyat *su gives [o'on (s →], u → o Formosan languages frequently add ? initially and finally, the -on is probably *-an. However, elsewhere in the paradigm we have [o?]. This, I think, forces us to conclude that the *yu found in the 2nd pl is not related to the *iu 2nd sg pronoun found outside Formosa. These are two cases where *yu occurs alone as a pronoun. The first, which we have already mentioned, as Tagabili yuh. Now this could be taken as good evidence that *yu occurred as a free-form pronoun. However, the Tagabili (Reid 1971) lpl excl is mih which must derive from *ami or *kami. This raises the possibility that Tagabili *yuh is from *kayu or some other form ending in yu. Therefore Tagabili cannot be said to provide good evidence for an independent *yu. The second situation is in certain Manobo languages (Harmon 1974) where it occurs as a clitic pronoun. As clitics are usually reduced forms of pronouns, it appears best to say that there is not enough evidence that *yu occurred as an independent 2nd pl pronoun. This leaves us with *mu and *muyu. Of these *mu is older both because it is unaffixed and because it occurs in the other paradigms we have reconstructed and *muyu does not.

For much the same reasons we would say that *su is the original PAN 2nd sg not *ka. *su occurs in the other paradigms we have reconstructed and *ka does not and it does not occur as a prefix on other pronouns as *ka does. The general trend in the non-Formosan AN languages for the ka prefix pronouns to replace the non-prefixed ones (in most languages *kita, *kami, *kamu have replaced *ita, *ami and *mu) argues that *su is the original form which was replaced by *kasu

in most languages. (Dempwolff (1934-38) who did not have Formosan data reconstructed *kaw from *kasu). It is an incontrovertible fact that *ka was a 2nd sg pronoun before the break up of PAN. The Tsou and Philippine evidence requires us to claim this.

The exact function of *ka in PAN does not appear to be determinable from the evidence available. The distinction between *ka and *su may have been the familiar one of deference/politeness but no present-day AN language reflects this. As already mentioned, Tsou is the only AN language which still has both as 2nd sg pronouns and in Tsou the conditioning for the use of one or the other is morphophonemic. This may have been the conditioning factor in PAN with *ka occurring after certain particles or in certain constructions and *su elsewhere. This could explain how *ka became a prefix. If it signalled 'second person singular' in certain environments, then it could probably generalise to signal 'second person' in those environments and so be added to *mu and *ita and from there be added to *ami by paradigmatic pressure so it lost its meaning. On the other hand *ka could have been some morpheme which by its own semantics was naturally associated with the second person (like a deictic). It could have been prefixed to the second person pronouns (which we take to include *ita lpl incl) in certain circumstances. As the singular is unmarked in the opposition plural/singular, the occurrence of *ka alone would presumably be interpreted as singular. This could explain the occurrence of *ka, *su and *kasu as opposed to *kita : *ita and *kamu : *mu. We should note that Amis has a series of ka versus non-ka pronouns. In Amis the non-ka forms are topic pronouns whereas the ka forms are nominative. However, it is unlikely that this is a reflection of the PAN differentiation. No other AN language has forms kako (from *ka + *aku) and kiso (from *ka + *i + *su). They are presumably innovations resulting from paradigmatic pressure of non-ka topic and ka nominatives in the first and second plural. This differentiation in the first and second plural must be an Amis innovation otherwise it would have left some traces elsewhere in AN. However, there are probably many other equally plausible or implausible scenarios explaining the relationship between *ka and *su. All we can do on the evidence available is reconstruct both. We must also reconstruct the form *kasu.

This means that we are suggesting the following first and second person nominative pronouns.

sg 1	aku/i-aku
2	su/i-su, ka, kasu/i-kasu
pl 1i	ita, kita/i-kita
1e	ami/i-ami, kami/i-kami
2	mu/i-mu, kamu/i-kamu

There was presumably some difference between the forms with i- and those without i-. *i is the common noun focus marker (i.e. it signals that the NP following it is by nominative). However, these pronouns without the i signal this anyway and so this cannot have been the function of the i pronouns in PAN. Yet their function must have been similar to that of the plain pronouns because they have replaced the plain form pronouns in many languages. Two languages which provide us with evidence as to the nature of the distinction between the *i and the plain pronouns are Atayal and Tagalog. Atayal has two sets of pronouns which Egerod (1966) calls nominative (corresponding to our nominative) and 'nominalised'. The two sets are the same except for the 2nd sg isu (nominalised) vs su (nominative), lpl incl ita? (nominalised) vs ta? (nominative) and 1st sg kuziŋʷkun (nominalised) vs sakuʷku? (nominative). The 1st sg forms are not relevant here but the difference between isu and su and ita? and ta? is of great interest. Egerod says that the nominalised forms occur "as complements

or in exposure". From the sentence examples in Egerod (1966) it appears that the nominalised pronouns function as sentence-level topics; reflexives; and contrastively. Much the same thing occurs in Tagalog for the 2nd sg which has two forms *ikaw* (from **ikasu*) and *ka* (from **ka*). *ikaw* functions as a sentence-level topic or contrastively while *ka* is used elsewhere.

There can be little doubt from these facts, I think, that the **i* pronouns must have functioned this way in PAN. The **i* pronouns are not the sort of phenomena which would arise independently in two languages. The functions of the **i* pronouns follow reasonably from the semantics of a combination of the focussed forms of the pronouns and the focus marker **i*. The function of **i* is to tell the hearer that the nominal following it is the centre of attention for the clause. Now if we already have special pronouns forms which convey this information, a combination of the two forms would surely indicate that there was something special about this pronoun as centre of attention for the clause. Contrastiveness is the most obvious example of this. Sentence-level topics and reflexives are also good examples. In each case there is something unusual and all three interrelate. Sentence-level topics and reflexives can be used contrastively: 'I was the one who did it', 'I did it myself'.

From these functions the **i* pronouns could easily spread to take over the functions of the plain forms. The connection between sentence topics and nominatives is very strong. Nominatives are unmarked themes and topics introduce new and contrastive themes. The contrastive use is quite likely to be devalued over time by overuse (consider 'terrible' and 'awful' in English). The use of these forms as sentence-level topics would also explain the forms in Kn and Saaroa. In these languages the *i* form pronouns are free form pronouns, and the plain forms are bound. The *i* form pronouns would be most unlikely to develop into clitics because of their functions whereas the plain forms could easily do so. One thing should be noted about the **i* forms. There is no evidence for a form **i-ka* which is another argument that **ka* is a later development as a 2nd sg and that **su* is the original pronoun.

This leaves us with the third person pronouns to reconstruct. Blust has reconstructed **t'i-ia*. This form is reflected by Atayal *hia*[?], Saisyat *siya*, Tondano *sia* and so on in many languages. Many other AN languages such as Malay, Ibanag, Nias and Gayo reflect simply **ia*. Now this pronoun is fairly easily analysable into a combination of the common noun focus marker **i* and the demonstrative **a*. **i* of course, occurs on all the other pronouns. Now this argues that PAN did not really have third person pronouns but had instead deictics which could be used to refer to entities not covered by the first and second person pronouns. The reanalysis of the demonstrative as a pronoun must have taken place in PAN. The reanalysis, however, was not a simple reanalysis of the demonstrative **a*, the form reanalysed was the compound **i* + **a*, a focussed and possibly nominalised form (meaning 'that one').

The question arises as to why the third person is the only person with a variant in **t'i*. Now **t'i* is usually reconstructed as the personal noun focus marker with **i* being reconstructed as the common-noun focus marker (i.e. when these morphemes occur before a noun they indicate whether that noun is common or proper and that it is by nominative). However, there are good arguments for reconstructing **a* as the focus marker for common nouns. It occurs as such in Paiwan, Palauan, Tagalog and other languages. Until now we have been using Reid's (1978) reconstruction of the personal noun FM as **t'i*, the common noun FM as **i* and **a* as a demonstrative. However, if these were the PAN forms why do the first and second person pronouns which are inherently proper take the common-

noun focus marker *i, and there is no evidence with the possible exceptions of Atayal and Bunun that they took *t'i.

Now, it is a well-known fact that pronouns tend to preserve older forms. The fact that *i is prefixed to the first and second person pronouns in combination with the fact that a number of languages reflect *a as the common noun FM makes it seem quite probable that at some stage of PAN *i was the proper noun FM with *a being the common noun FM. However, there are a number of facts which make this analysis problematic. Firstly there is the fact that *t'i can be reconstructed as the proper noun FM in PAN. It is reflected as such in Formosa, in the Philippines, Borneo and Madagascar. However, *i occurs as the proper FM in Gaddang, Umiray-Dumagat and then in Kapampangan which is in a different Philippine subgroup. But it also occurs as the common FM in many Philippine languages and Reid reconstructs it as such for PPh. Secondly Reid reconstructs *a as the PPh ligature (the ligature in AN is a marker which occurs in NPs to show that the head N and its attribute form an NP). It has this function in a number of Formosan languages such as Kavalan and Thao where it is not the common FM. However, in Palauan a is the article and not the ligature. Furthermore, why should one combine *i and *a to form a third person pronoun if they were simply the personal and common versions of the FM. It makes more sense for a combination of *i as a proper FM with *a as a demonstrative (Reid's proposal) to function as a third person pronoun (as we pointed out earlier with a meaning something like 'that one').

There is another area which needs considering before a solution to these problems is proposed. This is, of course, the third person plural. This has been reconstructed as *t'i-da. This form has generally been accepted as the PAN 3rd pl. However, there are reasons to dispute this assignment. The only possible cognate for *t'i-da as a 3rd pl in Taiwan is Atayal hga[?]. The development d → g is otherwise unknown in Atayal or AN. It is phonologically unlikely and so cannot be explained as some natural phonological change occurring sporadically in the pronouns. Therefore we must reject the Atayal form as non-cognate.

The obvious Taiwanese cognate for *t'i-da is the Amis 3rd sg tsira. This immediately raises the question of whether *t'i-da was a singular or plural and if it was a singular what *ia ~ *t'i-a was. There is another fact which is relevant in deciding these questions. This is the lack of third person pronouns in a number of Taiwanese languages. Bunun and Rukai have no third person pronouns, only deictics. Tsuchida (1976) does not give third person pronouns for Tsou, Kananabu and Saaroa. The forms of the third person pronouns in Paiwan, Pazeh, Thao and Kavalan are unrelatable to one another or anything else in AN. The Atayal and Amis third plurals are unrelatable to anything else in AN. In Saisyat and Sediq, *t'i-a is the root for both singular and plural.

There are two possible explanations for this state of affairs. Either the third person pronouns have been lost in many Taiwanese languages or they have been innovated as pronouns since PAN. The evidence clearly favours innovation. Accepting that the third person pronouns are innovations helps us with finding a solution to the problem with *i and *t'i, raised earlier. The best solution to that problem seems to be that *t'i is a PAN innovation. In pre-PAN times there was no proper vs common distinction in the construction markers, *i was the only FM. This explains why it occurs on the pronouns, which are, as we have already pointed out, inherently proper. Some time before the break-up of PAN *t'i was innovated as the proper FM. This naturally led to the creation of a *t'i-a form in competition with the already existing *i-a. We can propose a

plausible origin for *ia, *t'i-a and *t'i-da. We have suggested that *a was a PAN or pre-PAN demonstrative and that the focussed form *ia would be a natural choice for a morpheme to refer to the third person entities. Once the proper marker *t'i was innovated a form *t'i-a would naturally arise in competition with *ia.

The origin of *t'i-da is harder to determine. The *da could come from two sources. One is the da plural marker which occurs in the Philippines: *t'i-da could be simply 'those plural ones'. However, the occurrence of *t'i-da as a singular in Amis causes problems. The development of a 3rd pl into a 3rd sg is not something which is known to have occurred in the development of any other language. Another possibility is that *da was a deictic element like *a. A deictic formative da occurs in Hanunoo; Bisayan has da? which means 'near you' and Paiwan has a deictic tsad'a meaning 'far away'. If these forms are cognate, they could represent developments of a *da deictic. However, without a cognate from a language which preserves the distinction between *d and *d', we cannot be certain that the reflexes in the languages above do not represent a form *d'a. Therefore, it appears that for the moment we will have to leave the origin of *da as uncertain.

The occurrence of *t'i-da in a smaller number of languages than *ia ~ *t'i-a supports the view that *t'i is an innovation. As *t'i-da is the rarest of the third person forms, it must have been innovated later. Its occurrence with *t'i only makes it clear that by the time *da came to be used pronominally, *t'i was the proper case marker.

The non-occurrence of an oblique -ən form for the third person and the non-occurrence of a bound genitive separate from the free *ni- genitive, also provide strong evidence for the innovation of the third person pronouns. The situation we are proposing for pre-PAN is that there was a deictic *a which could be used pronominally in its focussed form. Later *t'i was innovated as a proper case marker and a form *t'i-a started to compete with *ia. Later still, an element *da came to be used pronominally in its *t'i form.

This means that we are proposing the following pronoun system for PAN:

Category 1 (nominative)

sg 1	aku/i-aku
2	su/i-su, ka, kasu/i-kasu
pl li	ita, kita/i-kita
le	ami/i-ami, kami/i-kami
2	mu/i-mu, kamu/i-kamu

Category 2 (genitive)

	bound	free
sg 1	ku	ni-aku
2	su	ni-su
pl li	ta	ni-ita
le	mi	ni-ami
2	mu	ni-mu

Category 3 (oblique)

sg 1	akən
2	sun
pl li	itən
le	amən
2	mun

and forms *ia*, *t'i-a* and *t'i-da* (nominative) *ni-a* and *ni-da* (genitive) which referred to third person humans. It should be noted that *ka* prefixation must precede the innovation of **t'i* because the *ka* pronouns are still prefixed with *i*. This means that *ka* prefixation is very old, which could explain the problems we had in determining its semantics.

2.3. PAN construction markers and deictics

From our discussion so far, we can see that there were at least three nominal categories which must have had separate construction markers (CMs). These are the three categories which occurred in the pronouns. We have already seen that in pre-PAN, the CM for Category 1 was **i*. In later PAN, there were two markers **t'i* for proper nominals and **i* for common nominals. The most widespread CM for the second category is **ni* (found in Paiwan, Tagalog, Toba Batak, Malagasy and Fijian). Blust 1977 also provides evidence for an **i* CM which may have been a postconsonantal variant of *ni*. Reid provides supporting evidence for this from the Philippine languages (Reid 1979).

In the languages where a common vs proper distinction is made, **ni* is the proper CM. The best candidate for the corresponding common CM is **nu*. Reid 1978 provides evidence for both **nu* and a form **na*. Both forms are widely reflected in Austronesian. However, **na* could be regarded as a reduction of **nua* (the form which is found in Paiwan). This form can be analysed as **nu* plus the **a* we have already encountered.

If pre-PAN did not have a common vs proper distinction in the CMs for category 1 then it is most unlikely that it had them for category 2. The difference between the two categories is that the nominative proper CM was innovated whereas the common genitive CM was innovated. This difference is probably due to the fact that Actors and Possessors are usually human and so the undifferentiated genitive marker **ni* would probably become a proper CM when the proper vs common distinction innovated in the nominatives was extended to the genitives.

From our discussion of clause structure it will be obvious that there are at least two CMs in the oblique category. These are of course **i* and **akən*. **i* would have marked non-PrP temporals and locations; **akən* would have marked non-Prp instrumentals, benefactives and comitatives. Another locative marker must be reconstructed for PAN. This is **di* which is reconstructed by Reid (1978) for PPh. It also occurs in Toba Batak and Pazehe. The difference in function between **i* and **di* is unknown. It is not known what CM was used to cover non-focussed patients and goals. The Toba Batak evidence mentioned earlier suggests that they were possibly simply bare NPs. However, it is equally possible that they were covered by the second category CM as occurs in many languages such as Paiwan and Tagalog.

We have already mentioned that **a* occurs as the common FM in Paiwan, Tagalog and Palauan. However, there are a number of other facts about **a* which point to it being a demonstrative, even though it is not reflected as such in any AN language. Firstly, as already pointed out, **a* must have been a deictic for the semantics of **ia* and **t'ia* to make sense. Secondly, Reid (1978) reconstructs *a* as the PPh ligature. A ligature is a morpheme which signals that two items are bound together as an NP. Ligatures occur widely in the AN languages (see Foley 1976). It occurs as a ligature in Pazehe (Li 1978) as well. It occurs as a ligature and the common FM in Paiwan, and Kavalan (Li 1978). In

Paiwan *a also functions as the morpheme which marks that two constituents are in an equational relationship. In Palauan *a is the article but not the ligature. However, there are further complications to this. Blust (1974) has reconstructed a PMP ligature *ŋ(a). Some languages have this and others, such as Ivatan, Maranao, Bilaan and Yogad, have only *a. The only way to explain these developments is to posit *a as a demonstrative originally as Reid (1978) has done. The use of demonstratives as ligatures is well known in many languages of the world ('that man', 'the man that I saw' English relative pronouns are equivalent to AN ligatures). The development of demonstratives into articles is also equally well known.

The exact function of *a in PAN is uncertain. It was clearly a demonstrative (probably a distal demonstrative in pre-PAN). Reid (1978) suggests that *a was the PAN ligature and that the a articles found in Paiwan, Tagalog and Palauan arise from reductions of *na (the PAN distal demonstrative). This hypothesis seems to involve unnecessary complications when there is already a more acceptable source for these forms. It seems better to accept *a as the PAN ligature and a demonstrative.

There are two other demonstratives we can reconstruct for PAN. One, already mentioned, is the *na distal demonstrative reconstructed by Reid (1978). The other is *tsu, which occurs as both a proximal and distal demonstrative in various AN languages (tsu 'this' Paiwan, itu 'that' Indonesian). However, as *na is nearly always reflected as distal, it is perhaps better to reconstruct *tsu as proximal.

Therefore we have a PAN CM system

	common	personal
FM	i	t'i
Gen	nu	ni
Loc		i,di
Inst		akən

and two demonstratives *tsu and *na and a ligature/demonstrative *a.

3. SUBGROUPING THE AUSTRONESIAN LANGUAGES

3.1. The PMP subgroup

We will now consider whether there is enough evidence for us to say that the MP languages form a subgroup. We will consider the phonological and grammatical evidence separately, taking the phonological evidence first.

3.1.1. Phonological evidence

The MP languages all show the following mergers:

- a) *t and *ts merge to *t;
- b) *l and initial *ʎ to *l;
- c) *n and non-initial *ʎ to *n.

Change a) is not particularly convincing as it is the sort of change which could arise in a number of different languages independently. However, changes b) and c) which are a split merger of *ʎ, *l and *n to provide convincing

evidence for a PMP subgroup. A split merger is not the sort of thing which arises independently as a result of drift. Some languages occasionally show no reflex for initial *ɬ sporadically. The examples Dahl cites are *ɬaŋuy 'to swim', *ɬu(n)tuŋ 'mortar', *ɬabuq 'to anchor' and *ɬa(n)tak, *ɬitan 'to use a pestle' which show n reflexes in some of the languages of Sulawesi, the Philippines and Chamorro. However in three cases, ɬaŋuy, ɬu(n)tuŋ and ɬa(n)tak, the ɬ is usually followed by a nasal (in most languages the n in (n) is present). Furthermore sporadic alternations between n and l (which is the reflex of initial *ɬ outside Formosa) occur elsewhere in AN. TB has lanok 'fly' from *lalaŋ 'fly' and Tg has ?a-ninao? 'explain' from *linaw 'realise' (Dempwolff 1934-38). These five words also exhibit the normal l reflexes in many other MP languages such as Tagalog, Javanese and Malagasy. Therefore I do not think that these five words can be said to present a challenge to the split merger.

Another equally, if not more, convincing piece of evidence for a PMP subgroup is the innovation of *n'. This phoneme must be reconstructed for the non-Formosan languages, but as we have demonstrated it appears to be an innovation of these languages. As we have already mentioned, it appears that Dempwolff had doubts about the status of *n' as a PMP phoneme. But it occurred in environments such that it was not phonologically reducible to a palatal allophone of *n (e.g. *n'amuk 'mosquito', *pən'u 'turtle', *n'an'i 'sing'). There is, however, an explanation which would show how *n' was innovated in PMP. I think that a consideration of the phenomenon of nasal grade will provide a good explanation for the origin of *n'.

AN languages normally do not permit consonant clusters. However, most of them do have nasal/stop clusters. These clusters are called nasal grade. There are two types of nasal grade. One is produced by prefixing maŋ- or ŋ- to a stem; this is called nasal substitution and does not occur in Oceanic. The other type of nasal grade is called prenasalisation or simply nasal/stop clusters. In Oceanic prenasalisation occurs both initially and medially, elsewhere it occurs only medially. Neither type of nasal grade occurs in Formosa. The two types of nasal grade have quite different effects.

We will give an example of nasal substitution from Ivatan to show its effects.

maŋ + ili 'village' → maŋili
 maŋ + punu 'full' → mamunu
 maŋ + bətbət 'pluck hairs' → mambətbət
 maŋ + takaw 'steal' → manakaw
 maŋ + kali 'dig' → maŋali
 maŋ + sawud 'knit' → manawud
 maŋ + mata 'raw meat' → maŋmata

These effects occur wherever nasal substitution occurs. A combination maŋ and a vowel has no special effect; maŋ and a voiceless stop combine to give a nasal homorganic to the stop; maŋ and a voiced stop give a homorganic nasal stop cluster; maŋ and s (from *t') give n' in those languages which have palatals and n in those which do not; otherwise maŋ has no special effects. maŋ is an AF affix in the focus system.

Prenasalisation is quite different in its effects and does not appear to have any meaning attached to it. Prenasalisation, except in one case, always produces homorganic nasal/stop clusters. However, this one exception is of great importance historically. In Javanese and Malay the reflex of *n't' is ŋs (Jv laŋse, Ml laŋsai from *lan't'ay 'shade'). As Dahl argues this ŋs must represent an original ŋt' which must be the prenasalised form of *t'. As far as we

know there is no reason to assume that *t' was different from any other stop in relation to prenasalisation so one must assume that all the homorganic nasal/stop clusters were originally ŋ/stop clusters.

There is another curious fact about prenasalisation. In his reconstructions involving the nasal/stop clusters Dempwolff almost always has the nasal in brackets. The brackets indicate that the reflexes of the nasal are not predictable. A language may have a nasal/stop cluster or simply the stop. If we examine the cognate sets we find that one language may have a nasal/stop cluster, another a stop, and a third may have both. It is completely unpredictable. Take for example *ta(ŋ)kub 'to cover' and *da(n)daŋ 'heat'.

Tg	TB	Jv	Ml
takob	takkup	takup	taŋkub
daraŋ/dandaŋ	dadaŋ	dandaŋ	dandaŋ

Such examples are numerous. It is obvious that the presence of the nasal was optional. This means that at one time ŋ could be optionally infixes before the medial stop in a word. However Gonda (1943) attacks the view that you can call this ŋ an affix. Gonda argues that you cannot call something an infix if there is no meaning reconstructible for it. It is true that there is no meaning reconstructible for ŋ. He shows that in many languages, both AN and non-AN, nasals and laterals are infixes to words to make them more expressive and emphatic. In AN *-al- and *-aŋ- occur as infixes with no reconstructible meaning. Therefore Gonda argues that we should reconstruct *-ŋ- as a strek form (a phoneme or group of phonemes added to words to make them more intensive). However, Latta (1978) argues against this stating that merely because one cannot reconstruct the semantics of an affix is not an adequate reason for denying that it was an affix. Therefore it appears that the status of *-ŋ- is somewhat uncertain.

Though nasal substitution and prenasalisation have different effects they appear to be related. From the variation maŋ-~ŋ-, we can suggest that maŋ- could be broken up into ma and ŋ. This ŋ would be the same one found in prenasalisation. The affix maŋ- which is structurally similar to maŋ- can also be broken up in this way as there is evidence for a *-ŋ- infix. This strengthens the idea of splitting maŋ- into ma and ŋ. This is in accord with Dahl (1976:128).

Once we have accepted this relationship, a number of questions arise for consideration. The first is for what level of AN do we wish to reconstruct these processes. This naturally entails determining which languages the processes can be reconstructed for. Secondly, we must explain the difference in effects between the two. Thirdly, we must decide whether prenasalisation occurred originally both initially and medially as in Oceanic or medially only as elsewhere.

The answer to the first question is that we would wish both types of nasal grade for PMP. There are only two forms which provide evidence for prenasalisation in Taiwan. One is the word for star, PAN *bi(t)uqən, Pz bintun, Ss bintö'än, Bun bintoqan 'star' and *qaŋt'u, Bun maqantsuh 'smell of rotten tubers' (Zorc). As homorganic nasal/stop clusters are probably the commonest type of consonant clusters cross-linguistically, we would require a reasonable number of forms before we could exclude the possibility of these forms being due to other factors than prenasalisation. Two forms is certainly not nearly enough to do this. Furthermore, there is no evidence for *maŋ- in Taiwan. If we accept that nasal substitution and prenasalisation are related then we must accept that

nasal substitution is older. Nasal substitution causes much more complex assimilations than prenasalisation which argues strongly that it is older. Secondly, when η was added to ma to form a focus affix it presumably had a much more specific meaning than when it could be infixes in any word base which again means nasal substitution must be older.

Nasal substitution does not occur in Oceanic, but there are traces of it (Reid, personal communication). So one can regard the loss of nasal substitution in Oceanic as being part of the general loss of the focus system. The Taiwanese languages, on the other hand, show no sign of having lost the focus system and no good reason can be proposed for the loss of $ma\eta\text{-}\eta$. Therefore we will regard $ma\eta\text{-}\eta$ as a PMP innovation.

We will assign prenasalisation to PMP. A number of northern Philippine languages such as Ivatan and Bontok do not show any evidence of prenasalisation, neither does Kelabit in Borneo (Blust, personal communication). However, as these languages on other evidence all belong to the PMP subgroup we will, for the moment, regard the absence of prenasalisation as being due to loss. It is possible that the absence of prenasalisation in a number of northern Philippine languages could indicate that prenasalisation was not innovated until later. As we have said, all the evidence indicates that nasal substitution is an older process than prenasalisation and this is the reason for their different effects.

I would say that on the evidence available that it is best to reconstruct prenasalisation medially only. The prenasalised stops which occur initially in Oceanic are probably better regarded as the trace of nasal substitution with η .

Once we accept that nasal substitution is a process innovated at PMP level then we will obtain a perfectly plausible origin for $*n'$. This phoneme would be produced by $ma\eta$ - prefixation to any stem starting with $*t'$. This would produce $*n'$ in all sorts of contrastive environments and would of course be completely undetermined phonologically. This would provide PMP with a $*n'$ phoneme which could then occur in new words added to the lexicon and spread into some old ones.

Therefore, from the phonological data we have the following phenomena which provide conclusive evidence for a PMP subgroup

- a) the split merger of $*n$, $*\eta$ and $*l$
- b) the creation of $*n'$ by nasal substitution.

We can also add the innovation of prenasalisation to these two pieces of evidence. However, as we indicated earlier, there are some doubts about what subgroup level this should be assigned to so it is not as strong a piece of evidence as the other two.

3.1.2. Grammatical evidence

The grammatical evidence for a PMP subgroup is not as extensive as the phonological evidence. One would have to reconstruct a grammatical system almost the same as the one we have reconstructed simply on the basis of the MP languages. The only language from Formosa which adds something to our reconstruction is Tsou and the Tsou evidence is basically confirmatory.

This is not say, however, that there is no evidence of grammatical innovations in a PMP subgroup. Considering the pronouns first, we find three excellent pieces of evidence for a PMP subgroup. The first of these is the

innovation of *t'i-da as a third plural. In Taiwan *t'i-da occurs only in Amis and there it occurs as a singular. The other two pieces of evidence are provided by Blust (1977). One is the use of *mu as a 2sg gen which does not occur in Taiwan where it is always a 2nd pl. As Blust points out, this is presumably due to politeness and deference restrictions. The third piece of evidence is provided by the oblique pronouns. Outside Taiwan the lpl excl *itən is reflected as *atən apparently under the influence of the lpl excl *amən. These three pieces of evidence provide more conclusive evidence for a PMP subgroup.

Clause structure provides us with one piece of evidence for a PMP subgroup. This is the innovation of *maŋ-ʌŋ- in the focus system. Blust (1974) provides us with evidence from the construction markers for a PMP subgroup. This is the ligature *ŋ(a). We have reconstructed *a as the PAN ligature and therefore *ŋ(a) must be an innovation.

The final area where one would like to have some evidence for a PMP subgroup is from the lexicon. However, as no studies have been done in this area, we have no evidence. With at least some words we know that they must be PMP lexical additions. These are words with *n' such as *n'amuk 'mosquito', *n'an'i 'sing' and *pən'u 'sea turtle' all of which are fairly basic and occur throughout the MP languages. I suspect that further studies would show much vocabulary which occurs exclusively in the MP languages.

From the evidence presented here, I think that one can conclude that a PMP subgroup of PAN did exist.

3.2. The grouping of the Formosan languages

Now that we have established that MP is a subgroup of AN, we naturally want to know how the Formosan languages group among themselves. The following subgroups have been proposed for the Formosan languages:

- a) Atayalic consisting of Atayal and Sediq;
- b) Tsouic consisting of Tsou, Kanakanabu and Saaroa;
- c) Favorlang-Pazeh-Saisyat;
- d) Amis-Bunun-Siraya-Kavalan;

with the rest being ungrouped.

The Atayalic group seems reasonably well supported by the evidence. Sediq and Atayal appear to have separated fairly recently. Sediq and Atayal have a form mu lsg gen which must be a common innovation since it is not found elsewhere in AN. They also share locative forms of 1st and 2nd sg kənan (s) knan (At) 1st sg, sunan (s+At) 2nd sg. Asai (1953) does not give full paradigms for any other pronouns and so we do not know if Sediq has locative forms for the other pronouns. However, on the basis of the first and second sg forms in Sediq it is highly probable that it does. As we have already mentioned, these locatives are an Atayalic innovation. They both have the PF marker as un indicating a common development of *ə to u. This is all that can be gleaned from the available sources but it appears to be good enough to say that Atayal and Sediq form a subgroup. I suspect that with more information there would be more evidence. Asai's Sediq grammar is rather fragmentary as an overall description of a language.

The evidence for a Tsouic subgroup is much less convincing than that for Atayalic. The phonological innovations which the three Tsouic languages share

can be summarised as:

- a) a merger of *ts and *d and *d' to ts in Kn and ts/s in Saar and Ts;
- b) development of final *an to a (all VC → VCV).

The merger of *ts and *d to *ts is a good argument for subgrouping the three languages because it is such an unusual merger. The development of *an to a is also a reasonable subgrouping argument because all other word-final groups VC acquire a supporting epenthetic vowel in Tsouic.

Grammatically the languages all show a tendency to merge PF and LF. In Tsou -a is the PF marker and in Saaroa it is the optional PF marker. In Saaroa the LF marker is -a(na) and in Kanakanabu it is -a in perf LF. In Kanakanabu imperfective LF and PF are both marked by -ənə. In Kanakanabu perfective PF is marked by -ə. Tsou and Saaroa have both replaced *-ən with -a (as we pointed out earlier this -a can come from PF *-a or LF *-an). On the other hand, Kanakanabu appears to have extended the use of *ən, at least in the imperfective. Therefore we have Tsou-Saaroa grouping against Kn. Kn and Saar merge *ɖ with *d, *d' and *ts whereas Tsou has an r reflex for *ɖ.

Most scholars split Tsouic into Tsou and Kanakanabu-Saaroa. However, there does not appear to be any evidence for this grouping. Tsuchida does not cite any grammatical innovations for a Kanakanabu-Saaroa subgroup. There is a lot of shared vocabulary between Kanakanabu and Saaroa and indeed among all three languages but this does not provide that they form a subgroup. Ferrell (1969:36-39) makes it clear that the Tsouic languages form a cultural group. This would undoubtedly lead to a great deal of borrowing between the languages. It appears therefore that there are considerable doubts about the existence of a Tsouic group.

Tsuchida (1976) suggests that Rukai groups with Tsouic. The evidence he offers is that the Maga and Mautauran dialects of Rukai have the same final *an to a development as in Tsouic. Like Tsouic, Rukai adds epenthetic vowels to consonant-final stems. These are the only two common innovations. However, the *an to a change does not occur in all Rukai dialects so it cannot be used as a subgrouping argument. The occurrence of the change *an to a in certain Rukai dialects (the ones adjacent to Tsouic) must lessen its value as a subgrouping argument for Tsouic and thereby increase the doubts about the Tsouic subgroup. The innovation of the epenthetic vowel is not a particularly convincing subgrouping argument because it occurs elsewhere in AN as in Sangirese and Palauan. Furthermore it is a fairly natural result of pronouncing final consonants with release.

There is one piece of evidence that Rukai may have subgrouped in Paiwan. In an archaic text concerning headhunting collected by Li, the 2nd pl Category 1 pronoun is mon. This obviously derives from *mun, the PAN Category 3 pronoun. Now Paiwan is the only other AN language which has Category 1 pronouns from PAN Category 3 pronouns. However this one pronoun form is not enough to group Rukai and Paiwan. However if Rukai ever grouped with anything then all traces of this have since been lost by its own innovations as can be seen from the pronoun paradigm (see pronoun paradigms, section 2) and the verb system (see Li 1973:156).

The next subgroup which has been proposed for Formosa is Favorlang, Paze, Saisyat (Marsh 1977). The evidence for the subgroup is that all three languages merge *dz (which we do not accept) and *ɖ. In Paze and Saisyat these two phonemes merge with *d as well. Unfortunately, there is no other evidence for a subgroup that I have been able to find. The Paze and Saisyat pronoun paradigms (set out in section 2) show no evidence of a period of common development. There is no

paradigm available for Favorlang. Nor does it appear that they share any innovations in syntax or CMs. The Pazeh and Saisyat CMs are (Li 1978):

	Pazeh	Saisyat
Topic	ka	∅
Nominative	ki	∅
Accusative	a, u	ka
Genitive	ni	noka, na, ni (pers)
Locative	di	ray, kah
Instrumental	∅	noka
Comitative	∅	ni

I have not altered Li's terminology because the functions of the CMs are not relevant and because I do not have enough information to determine their functions. What is relevant is that the forms of the CMs appear to be unrelated except for the *ni* genitive which is PAN. Again I have not been able to obtain any Favorlang forms. From Marsh's description, Pazeh, Saisyat and Favorlang verbal systems do not appear to show any common innovations. The Favorlang focus affixes are *m* AF, *-en* PF, *-anvayan* LF, *i-* IF, *-in-* perf; in Pazeh these are *mv-* (*v* homorganic to first vowel in stem) AF, *-un* PF, *-an* LF, *si:-* IF and *-in-* perf; and in Saisyat these are *-om-*, *m-* AF, *-in* PF, *-an anai* LF, *si-* IF and *-in-* past. They all have *pa* causative. In all the languages perfective PF constructions have only *-in-* and no **-ən*. These things are all reconstructible for PAN and naturally therefore cannot be subgrouping arguments. From the information available at the moment, there is only one innovation, a merger, to support the existence of a Pazeh-Saisyat-Favorlang subgroup. Therefore we must regard the status of this subgroup like that of Tsouic as questionable.

Finally we will consider the Amis-Bunun-Kavalan-Siraya (Ferrell 1979). These languages all merge **t* and **ts* to *t*. Amis and Kavalan have merged medial **g'* and **n* to *n*. Bunun and Kavalan have merged **n* and **ɲ* to *n*. Amis has merged **ɟ* with **t* and **ts*. Amis has an accusative CM *tu ... (an)* (Reid 1971). The non-personal accusative in Kavalan is *tu* (Li 1978). These two CMs must be cognate. This CM also occurs in Thao (Li 1978). Thao merges **t* and **ɟ* to *t* just like Amis. Thao shares the irregular genitives *nak* and *nam* with Bunun. Thao and Kavalan have **a* as a ligature with the following variants, *a* after consonants, *wa* after back vowels and *ya* after front vowels. This suggests some sort of common development since PAN. Now these facts all point towards Thao having some sort of relationship with the subgroup we have proposed though it lacks the **t* and **ts* merger which defines the group. However, if we examine the mergers which we have been considering we find the following facts.

- *n* and **ɲ* merged in *n* in Kv, Bun and Kn.
- *t* and **ts* merged in Am, Bu, Kv, Sir.
- *d* and **ɟ* merged in Bun, Th, Kn, Pz, Ss, At, Se, Sar.
- *t* and **ɟ* merged in Am, Th.
- *d* and **ts* merged in Th, Ts, Kn, Sar.

All these mergers involve continuous areas except for the Amis and Thao merger of **t* and **ɟ* to *t*. The impression one gets when examining these mergers is that a wave model would be a far more accurate model of the situation than a tree diagram. The only mergers which are really convincing as subgrouping evidence are Thao and Amis **t* and **ɟ* to *t* and the merger of **ts* and **d* to *ts* in the Tsouic languages. These mergers are unusual and therefore argue for some common period of development. The Favorlang-Pazeh-Saisyat group looks weaker than ever if we consider the fact that Pazeh-Saisyat and Bunun all merge **d*,

and *ɖ to d (or r in Saisyat). Amis and Thao share the unusual *t and *ɖ merger to t, which cannot be areal, and the accusative marker tu. However this last thing occurs elsewhere. It seems that we will be able to subgroup languages only on the basis of a large number of shared similarities. At the moment it seems that there is a case for subgrouping Atayal and Sediq into Atayalic and much less convincing arguments for subgrouping Saaroa, Kanakanabu and Tsou into Tsouic and Amis and Thao.

The problems we have in subgrouping the Formosan languages can have two explanations. One is that they all form one subgroup, or at least the non-Atayalic-Tsouic languages form a subgroup. The only piece of evidence I know of for a Formosan subgroup is the fact that *d and *d' are always reflected the same way in the languages of Formosa. If one examines this merger more closely, one finds that this does not provide any evidence for a subgroup. In Paiwan *d and *d' merge to d' (i.e. d becomes d'). In Siraya *d and *d' merge to d with the merger going in completely the opposite direction. In Amis *d, *d', *l, *ɣ all merge to an alveolar tap so we have no idea whether *d and *d' ever were merged by themselves. In Kanakanabu *ts, *d, *d', *dz and *ɖ all merge to ts so again we have no idea whether *d and *d' ever had an independent merger. The same can be said for Pazeh and Saisyat where *d', *d, and *ɖ all become d and r respectively. Apart from these problems this merger would not uniquely identify the Formosan languages anyway. The languages of the Philippines all merge *d and *d' as well. Therefore we would have to subgroup the Philippines and Formosa which goes against the much better evidence we have presented for PMP. There does not seem to be convincing or even vaguely suggestive evidence that there is a Formosan subgroup. We will consider an alternative solution in the next section.

3.3. The relationship between the PMP subgroup and the Formosan languages

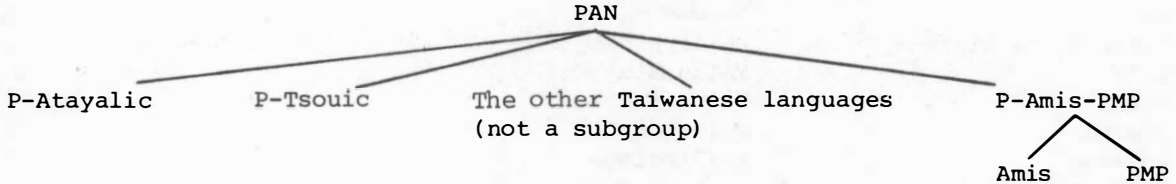
The question now is whether PMP is a first-order subgroup of PAN, whether it groups with one of the Formosan subgroups we have suggested or one of the ungrouped Formosan languages. PMP does not group with Atayalic or Tsouic. It shares no innovations with these subgroups. Nor does it appear to share any innovations in common with most of the ungrouped languages.

However, it does share the mergers of *t and *ts to t with Amis, Bunun, Siraya and Kuvulan. Of these languages, PMP appears to have the most similarities with Amis. Amis has the ka pronouns which the other three languages do not. The ka pronouns are the dominant forms in PMP; the non-ka pronouns occur only rarely in some Philippine languages. Amis is the only Taiwanese language with a reflex *t'i-da which we have shown to be an innovation. This common innovation of *t'i-da is an excellent argument for subgrouping PMP and Amis. However, there does not appear to be any argument for assigning any other language to this Amis PMP subgroup.

Therefore we have Amis PMP and Atayalic with a doubtful Tsouic subgroup and the rest of the Taiwanese languages being subgrouped. This interpretation of Austronesian subgrouping can have only one meaning historically. This is that Taiwan is the homeland of the Austronesians. High internal diversity among the Taiwanese languages combined with the fact that all the non-Taiwanese languages form a subgroup which can in turn be grouped with Amis, compel this conclusion (Sapir 1949b, Bloomfield 1933). This conclusion is in line with the archaeological and anthropological evidence (Bellwood 1978).

This would explain why it is so difficult to subgroup the Taiwanese languages. If Taiwan is the homeland then the Taiwanese languages have been interacting for at least 5,000 years (the Malayo-Polynesian departure from Taiwan must go back 5,000 years). Greater than 5,000 years of interaction between related languages is enough time to obliterate any original subgroup in Taiwan. As we have said, if we have Taiwan as the homeland with one group of people the Malayo-Polynesians leaving Taiwan to settle the rest of Austronesia then we would expect the Taiwanese languages to be as different from one another as they are from PMP.

Therefore we propose the following PAN family tree:



NOTE

List of languages mentioned, with abbreviations as used in the text.

Language	Location	Abbreviation
Acehnese	Sumatra	
Amblau	Indonesia	
Amis	Taiwan	Am
Atayal	Taiwan	At
Bikol	Philippines	Bk
Bisayan	Philippines	
Buginese	Sulawesi	
Bunun	Taiwan	Bun
Cebuano	Philippines	
Fijian	Fiji	Fj
Gayo	Sumatra	
Hiligaynon Bisayan	Philippines	Hl Bs
Ibanag	Philippines	
Ivatan	Islands between Philippines and Formosa	Iv
Javanese	Java	Jv
Kanakanabu	Taiwan	Kn
Kavalan	Taiwan	Kv
Kelabit	Borneo	Kel
Madurese	Madura	Md
Magori	Papua	
Malagasy	Madagascar	Mal
Malay	Malaya	Ml

Language	Location	Abbreviation
Ngadju Dayak	Borneo	NgD
Nias	Islands off Sumatra	
Okolod Murut	Sabah	
Paiwan	Taiwan	Pw
Palauan	Palau Islands	Pl
Pazeh	Taiwan	Pz
Puyuma	Taiwan	Pu
Rukai	Taiwan	Ruk
Saaroa	Taiwan	Sar
Saisyat	Taiwan	Ss
Samar Leyte Bisayan	Philippines	SL Bs
Saʔa	Melanesia	Sa
Sediq	Taiwan	Se
Siraya	Taiwan	Sir
Subanon	Philippines	
Sundanese	Western Java	
Tagalog	Southern Luzon (Philippines)	Tg
Thao	Taiwan	Th
Toba Batak	Sumatra	TB
Tongan	Tonga	To
Tsou	Taiwan	Ts
Western Bukidnon Manobo	Philippines	WBM

Language Groups

Malayo-Polynesian - the non-Formosan Austronesian languages	MP
Oceanic - the languages of Micronesia Melanesia and Polynesia	OC
Philippine - the languages of the Philippines	Ph

Proto-Languages

Proto-Amblau	Indonesian	
Proto-Austronesian		PAN
Proto-Malayo-Polynesian		PMP
Proto-Oceanic		POC
Proto-Philippine		PPh

BIBLIOGRAPHY

ANDERSEN, Henning

- 1973 Abductive and deductive change. *Language* 49:765-793.

ANTTILA, Raimo

- 1972 *An introduction to historical and comparative linguistics*. New York: Macmillan.

ARMS, David G.

- 1974 Transitivity in Standard Fijian. Ph.D. dissertation, University of Michigan.

ASAI, Erin

- 1936 *A study of the Yami language, an Indonesian language spoken on Botel Tobago island*. Leiden: Burgersidijk & Niermans.

- 1953 *The Sedik language of Formosa*. Kanazawa, Japan: Cercle Linguistique de Kanazawa, Kanazawa University.

BELLWOOD, Peter

- 1978 *Man's conquest of the Pacific*. New York: Oxford University Press.

BLOOMFIELD, Leonard

- 1933 *Language*. New York: Henry Holt; London: Allen & Unwin.

BLUST, Robert A.

- 1974 Proto-Austronesian syntax: the first step. *Oceanic Linguistics* 13:1-15.

- 1976 Review of Otto Chr. Dahl *Proto-Austronesian*. *Language* 52:221-237.

- 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 9/2:1-15.

- 1978 *The Proto-Oceanic palatals*. Memoir 43. Auckland: The Polynesian Society.

CHEN, Matthew Y. and William S.-Y. WANG

- 1975 Sound change: actuation and implementation. *Language* 51/2:255-281.

CHEN, Teresa M.

- 1982 Verbal construction and verbal classification in Nataoran-Amis. Ph.D. dissertation, University of Hawaii.

DAHL, Otto Chr.

- 1976 *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph 15. Lund: Studentlitteratur.

DARDJOWIDJOJO, Soenjono

- 1978 Nominal derivation in Indonesian. In: Wurm and Carrington, eds 1978: 503-528.

DELBRÜCK, Berthold

- 1884 *Einleitung in das Sprachstudium*. Zweite Auflage. Leipzig: Breitkopf und Härtel.

DEMPWOLFF, Otto

- 1934- *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols.
1938 Berlin: Reimer.

DIXON, R.M.W.

- 1979 Ergativity. *Language* 55:59-138.

DYEN, Isidore

- 1951 Proto Malayo-Polynesian *Z. *Language* 27:534-540.
1956 The Ngaju-Dayak 'old speech stratum'. *Language* 32:83-87.
1974 The Proto-Austronesian enclitic genitive pronouns. *Oceanic Linguistics* 13:17-31.

EGEROD, Søren

- 1966 Word order and word classes in Atayal. *Language* 42:346-369.

EMENEAU, M.B.

- 1956 India as a linguistic area. *Language* 32:3-16.

FERRELL, Raleigh

- 1969 *Taiwan Aboriginal groups: problems in cultural and linguistic classification*. Academia Sinica Monograph 17. Taipei: Institute of Ethnology.
1979 Phonological subgrouping of Formosan languages. In: Naylor, ed. 1979:241-254.
1982 *Paiwan dictionary*. PL, C-73.

FOLEY, William A.

- 1976 Comparative syntax in Austronesian. Ph.D. dissertation, University of California, Berkeley.

GAUCHAT, Louis

- 1905 l'unité phonétique dans le patois d'une commune. *Aus romanischen Sprachen und Literaturen: Festschrift Heinrich Morf*, 175-232. Halle: Max Niemeyer.

GIVÓN, Talmy

- 1975 Topic, pronoun, and grammatical agreement. In: C.N. Li, ed. *Subject and topic*, 149-188. New York: Academic Press.

GONDA, J.

- 1943 Inwendige nasaal- en liquida-verbindingen in Indonesische talen. *Bijdragen tot de taal-, land- en Volkenkunde (BijdrTLV)* 101:141-206.

HARMON, Carol M.

- 1974 Reconstructions of Proto-Manobo pronouns and case marking particles. *WPLUH* 6/6:13-46.

HAUDRICOURT, André G.

- 1965 Problems of Austronesian comparative philology. *Lingua* 14:315-329.

HENDERSON, Eugénie J.A.

- 1965 The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15:400-434.

JENG, Heng-hsiung

- 1977 *Topic and focus in Bunun*. Academia Sinica Special Publication 72. Taipei: Institute of History and Philology.

JOSEPHS, Lewis

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

KEENAN, Edward L.

- 1976 Towards a universal definition of subject. In: C.N. Li, ed. *Subject and topic*, 303-333. New York: Academic Press.

KRISHNAMURTI, Bh.

- 1978 Areal and lexical diffusion of sound change: evidence from Dravidian. *Language* 54/1:1-20.

LABOV, William

- 1963 The social motivation of a sound change. *Word* 19:273-309.
- 1966 *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- 1972 The social stratification of r in New York City department stores. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- 1974 On the use of the present to explain the past. In: L. Heilmann, ed. *Proceedings of the Eleventh International Congress of Linguists... Bologna, 1972*, vol.2:825-851.

LATTA, Christian

- 1978 Internal reconstruction. Ph.D. dissertation, Ohio State University.

LI, Paul Jen-kuei

- 1973 *Rukai structure*. Academia Sinica Special Publication 64. Taipei: Institute of History and Philology.
- 1975 *Rukai texts*. Academia Sinica Special Publication 64/2. Taipei: Institute of History and Philology.
- 1978 The case marking systems of the four less-known Formosan languages. In: Wurm and Carrington, eds 1978:569-615.

MARSH, Mikell A.

- 1977 Favorlang-Pazeh-Saisiat: a putative Formosan subgroup. Ph.D. dissertation, Washington State University, Pullman.

MARTINET, André

1952 Function, structure, and sound change. *Word* 8:1-32.

NAYLOR, Paz Buenaventura

1975 Topic, focus, and emphasis in the Tagalog verbal clause. *Oceanic Linguistics* 14/1:12-79.

NAYLOR, Paz Buenaventura, ed.

1979 *Papers from the Second Eastern Conference on Austronesian languages*. Ann Arbor: University of Michigan.

PAWLEY, Andrew and Lawrence A. REID

1976 The evolution of transitive constructions in Austronesian. *WPLUH* 8/2:51-74.

REID, Lawrence A.

1971 *Philippine minor languages: wordlists and phonologies*. *Oceanic Linguistics* Special Publication 8. Honolulu: University of Hawaii Press.

1974 The pronominal systems of Proto-Cordilleran, Philippines. *WPLUH* 6/6:1-12.

1978 Problems in the reconstruction of Proto-Philippine construction markers. In: Wurm and Carrington, eds 1978:33-66.

1979 PAN genitive alternation: the Philippine evidence. *WPLUH* 11/2:45-55.

SAPIR, Edward

1949a The psychological reality of phonemes. In: D.G. Mandelbaum, ed. *Selected writings of Edward Sapir in language, culture and personality*, 46-60. Berkeley: University of California Press.

1949b Internal linguistic evidence suggestive of the northern origin of the Navaho. In: Mandelbaum, ed. 1949:213-224.

SCHACHTER, Paul

1977 Reference-related and role-related properties of subjects. In: Peter Cole and Jerrold M. Sadock, eds *Syntax and semantics*, vol.8: *Grammatical relations*, 279-306. New York: Academic Press.

SCHACHTER, Paul and Fe T. OTANES

1972 *Tagalog reference grammar*. Berkeley: University of California Press.

SCHUCHARDT, Hugo

1885 *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Robert Oppenheim. Reprinted and translated in Vennemann and Wilbur, eds 1972:1-37, 39-72.

SILVERSTEIN, Michael

1976 Hierarchy of features and ergativity. In: R.M.W. Dixon, ed. *Grammatical categories in Australian languages*, 112-171. New Jersey: Humanities Press; Canberra: Australian Institute of Aboriginal Studies.

SMITH, Neilson V.

- 1973 *The acquisition of phonology: a case study*. London: Cambridge University Press.

STURTEVANT, Edward H.

- 1917 *Linguistic change*. Chicago: University of Chicago Press.

TRUDGILL, Peter

- 1971 The social differentiation of English in Norwich. Ph.D. dissertation, University of Edinburgh.

TSUCHIDA, Shigeru

- 1976 *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph 5. Tokyo: Toyo Shuppan.

TWADDELL, William F.

- 1935 *On defining the phoneme*. Language Monograph 16. Baltimore.

VENNEMANN, Theo

- 1972 Phonetic analogy and conceptual analogy. In: Vennemann and Wilbur, eds 1972:181-204.

VENNEMANN, Theo and Terence H. WILBUR, eds

- 1972 *Schuchardt, the Neogrammarians, and the transformational theory of phonological change*. Linguistische Forschungen 26. Frankfurt: Athenäum.

WANG, William S.-Y.

- 1969 Competing changes as a cause of residue. *Language* 45/1:9-25.

WOLFF, John U.

- 1973 Verbal inflection in Proto-Austronesian. In: Andrew Gonzalez, ed. *Parangal kay Cecelio Lopez*, 71-91. *Philippine Journal of Linguistics* Special Monograph 4. Quezon City.

- 1974 Proto-Austronesian *r and *d. *Oceanic Linguistics* 13:77-121.

WURM, S.A. and Lois CARRINGTON, eds

- 1978 *Second International Conference on Austronesian Linguistics: proceedings*. PL, C-61. (2 fascicles). [SICAL].

ZORC, R. David

- 1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. PL, C-44.

TOWARDS A TYPOLOGICAL RECONSTRUCTION OF THE VERBAL SYSTEM IN PROTO-AUSTRONESIAN

J.C. Anceaux

In Austronesian linguistics comparative morphology has never had the attention paid to comparative phonology. There is nothing exceptional about this; it is – or has been – the case in the comparative linguistics of most of the other language families. The study of sound laws has to precede morphological comparison, because without them identification of cognate morphemes is not possible. Another reason why morphology has to wait for the results of the comparative study of sounds, is that historical phonology gives more reliable means for classification. Sound-changes, although showing some universal tendencies, have the advantage of being independent of meaning (at least in most cases) and, therefore, are unique in the combination in which they occur. For that reason, they are better indications for common or separate innovations. In morphology structural coherence and meaning may cause parallel developments. And insight in the classification is necessary for assigning a certain feature or change to a certain branch or to a proto-language. Finally, morphological comparison asks for a much deeper and more thorough knowledge: the data cannot be collected from wordlists and dictionaries and grammatical descriptions usually are a poor substitute for personal familiarity with the languages in question, as exceptions and irregularities may provide clues equally important as those found in rules and regular phenomena. This precludes fast progress, especially in the case of the Austronesian family with its great number and variety of languages.

As to the methods of comparative morphology, the most direct and 'classical' approach is that of *morpheme reconstruction*: grammatical morphemes are reconstructed in the same way as basic morphemes and words, their meaning and function are taken from what is common in meanings and functions of their reflexes in the various languages. But every change in a morphological system means restructuring of that system and may bring about change in the distribution of functions of its elements. This makes reconstruction of the meaning and function of grammatical morphemes a hazardous if not impossible enterprise.¹

A valuable complement is found in *typological reconstruction*: the various morphological structures are compared and a hypothesis is formed as to the system from which they developed. In this approach use is made of the fact that sometimes morphological distinctions persist or even return, though their markers

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, 101-110. *Pacific Linguistics*, C-75, 1982.

© J.C. Anceaux

are lost or replaced.² Typological comparison is mainly used for descriptive purposes, to draw conclusions as to functions in a certain language from what can be learned from other languages, and is closely connected with general studies of grammatical theory.³

Another complementary method is *internal reconstruction*: looking for indications for a former state of a language, mainly from inconsistencies in its present system.⁴ Very similar is the use of petrified forms and all other debris of a previous state of morphology. As the persistency of forms, even after the breakdown of a system, is the main principle, it makes a good complement to the typological approach.

Finally, it must be kept in mind that morphological comparison is impossible without constant reference to syntactical phenomena, as morphological elements are often replaced by syntactical constructions and vice versa.

Most of the monographic publications on AN morphology concentrate on verbal morphology.⁵ Of the latest ones which reached for at least an outline for reconstruction Wolff's study (1973) suffered from being built on materials from a limited number of languages and attributed the Philippine system too easily and without real proof to PAN, whereas the paper by Pawley and Reid (1976), though more thorough-going than anything else on the subject, leaves one question without a definite answer (the historical priority of either the Philippine or the Oceanic type of 'transitivising') and another problem untouched: the position of the so-called 'conjugated' verb forms in the historical development of the verbal system.

In the present paper we will try to do full justice to these conjugated forms, as in our opinion they play an important role, not only in most of the Western AN languages but also in the development of the Oceanic languages. Further we will show that certain phenomena in languages showing the Oceanic type clearly point to a solution of the priority question in the direction of the Philippine type being the most likely to be assigned to PAN, thus corroborating Wolff's rather aprioristic choice and Pawley and Reid's reasoned preference.

One of the most important questions one has to ask about the functional categories determining the character of the AN verbal system is: do the terms we are used to in western languages also apply here and are they sufficient? The most important question is: can all verbal forms be classified as either active or passive? It seems that this question was seldom explicitly posed and the answer too readily taken for granted.

Another matter which has to be pointed out beforehand is the possibility of *homomorphy*: two or more morphological functions or categories being marked by a morpheme of the same shape, like *s* as a marker of both plural and genitive of nouns in English. In several AN languages homomorphy occurs to such an extent that it has exercised a confusing influence on many descriptions. Historically it raises the question if the homomorphy indicates a morphological split, that is: was the morpheme originally just one, or has a merger taken place of what was a set of different morphemes originally (like **-en* and **-an* merging in Malay because of a regular sound-change).

The questions about the existence of passive forms, or whether certain verb forms had to be regarded as passive or not, was raised a long time ago⁶ and there will be ever more answers to it as more data become known and new languages taken into consideration.⁷ But these discussions have no direct bearing on the question whether a passive is reconstructible for PAN. Clark (1973) concluded that POC must have had an agentless passive. Passive forms implying unimportance of the

agent (e.g. Indonesian *ter-* forms) or not admitting the agent to be mentioned at all (e.g. Wolio *to-* forms) are of so general occurrence that it is hardly conceivable that such forms did not exist in PAN, although their formal markers show a number of different shapes. However, many languages show forms which are often also labelled 'passive' with which the agent is not left out or removed to the background. In many languages the constructions with these forms and the way they are used are such that they raise doubts as to their passive character. As it may be difficult to decide whether the agent or the goal is seen as the more important element, so that it seems that they are of equal importance, the greatest importance being given to the action itself, these forms may be regarded as representing something in between.⁸ It is exactly in this in-between area that we find the so-called 'conjugated forms'; therefore, we will devote some attention to them first.

What is meant by this term is: verb forms having an actor-marker, usually a prefix, incorporated. The distinctions made by these markers correspond to differences in the system of the personal pronouns, though not all distinctions of person and number found in the pronominal system need to be made or represented in the verbal markers. A great number of languages having conjugated forms and systems of their use were described by Haaksma (1933), however, without any comparative conclusions. Haaksma's book is far from exhaustive, as it was based on materials from published sources available at the time, that is almost fifty years ago. To the languages described by Haaksma as having conjugated forms many can be added now.⁹ A full-fledged system in which all persons and numbers are separately marked is found in Simalur, Nias, Mentawai, Busang, Mori, Napu, Sumba, Lamaholot, Roti, Kei, Buli, Weda, Biak, and the Yapen languages. The same, but without a distinction of number for the third person, is found in Aceh, Makasar, Sa'dan, Timor, and Minangkabau, without the distinction of number for the second person in Solor, and without number distinction in both the second and the third person in Buginese, Bare'e, Wolio, Bima, Kupang, and Old Javanese.

Some systems show one or two deviating forms. Malay and Lampung, for instance, have a third person form with prefix *di-* and the possessive suffix for the third person. In Toba and Mandailing we find *di-* forms with possessive suffix for both the second and third person. In Gayo and Mandailing the situation is the same, but there the prefix is *i-*. Dairi has *i-* and possessive suffix for the second person, the third person having the prefix *i-* only. These forms have played an important role in the discussion about the origin and function of verb forms in AN languages. On the one hand the prefixes look like ordinary actor-markers and are found as third person prefixes (without any suffix) in several languages (Dairi, Mentawai, Biak). On the other hand there are quite a few languages having passive forms with a comparative affix and a possessive suffix marking the agent and underlining the nominal character of these forms.¹¹ Some languages — e.g. Malay — show both forms; Malay has *di-* forms used as passive or goal-focussed forms with or without the agent explicitly mentioned and *di-* forms with the third person possessive suffix *-nya*. The most likely hypothesis for this double character of the *i-/di-* forms is that these prefixes were used for marking passive participles and also as third person actor markers and that this homomorphy caused the disappearance of one of the two forms in most languages and restructuring of the original system in others. A quite different position was taken by Wils (1952) who regarded the actor-suffix forms as proof that all conjugated forms are in essence possessive constructions, constituting a possessive (though really verbal) flexion, different from both active and passive. This more or less forced him to explain all actor prefixes

as of possessive origin, a way of reasoning which finds very little support in the structures and forms of the languages in question. Following in C.C. Uhlenbeck's (his teacher's) footsteps, Wils further looked for a connection between this possessive flexion and the alienable/inalienable opposition in possession marking, at least in those languages having a complete set of conjugated forms.¹² This last hypothesis is interesting enough but asks for further research. What Wils said about the conjugated forms being neither active nor passive deserves our full attention. First it must be said that this was not a brand new idea: more or less the same had been said by Berg (1937:101-109). The best evidence for such a position, however, is found in those languages which have not only a set of actor-prefixes but also a series of suffixes or enclitical forms referring to the goal of the action. In Wolio the use of such double-reference forms is obligatory in transitive constructions, also in those which have goal and/or agent explicitly mentioned. Which of the two - goal or agent - is topic in a sentence is indicated by word order and intonation:

o anaana akamatea o buea *'the child sees the crocodile'* (focus: child)
 o buea akamatea o anaana (with different intonation, same meaning, but
 focus: the crocodile)
 akamatea o buea o anaana (same meaning, no focus on either the child or
 the crocodile).

This makes clear that the double-reference forms can not be simply listed as either active or passive. The same is true for corresponding forms in other languages, although some of them show a tendency to avoid these forms if the agent is in focus in which case the active forms are used. The same can be said about the actor-prefixed forms in many languages.¹³ In Makasar and Bugis conjugated forms of intransitive verbs may be used if an adverbial phrase is focus or topic: Mak. iyami kunumera *'this I cry (about)'*; muko kulampa *'tomorrow I go'*; apa nuboya nubattu mae? *'what are you looking for (that) you come here?'*. (Esser 1929:165-166). This is the more remarkable, as these languages do not use actor-prefixes with intransitive verbs in any other context (see below).

The question, whether the post-verbal forms indicating the goal go as far back historically as the actor-prefixes, cannot be answered yet. The latter ones are found in many more languages than the former. Moreover, the goal-markers are less fully incorporated in the word-form than the actor-prefixes. In Wolio the goal-markers cannot be denied the status of suffixes, but some aspect-markers, whose tie with the nuclear morpheme is looser, may be inserted between goal-suffix and verb-stem. For Uma the goal-markers are explicitly called enclitics by Esser (1964:36). This clitical character is also apparent in Bugis and Makasar.

Many languages (especially found in Eastern Indonesia) have conjugated forms also for the intransitive verbs.¹⁴ Some languages, Makasar, Bugis, and Uma among them, have the enclitics indicating goal with transitive verbs also for indicating the 'agent' of intransitive verbs, so that the result may well be regarded as an ergative construction. Still a construction of agent-marker+verb+goal-indicator (suffix, clitic, or pronoun) is a good hypothesis for a historical prototype, not only for the 'conjugating' languages but also for the Oceanic languages not having any 'conjugation'. It is quite conceivable that a full pronoun referring to the actor was added whenever a pronominal actor had some stress, as is still the case in Buli: iseli *'I eat'*, ya iseli *'I eat (the others don't)'*. In such cases the prefix becomes redundant and might be dropped. If the same happens to both actor-prefixes and post-verbal goal-indicators, the result is exactly what we find in languages like Fijian: eratou sa yādra na

cauravou (*they asp.-marker get-up the young-man*) '*the young men got up*'; au a raici iratou na gone (*I tense-marker see them child*) '*I saw the children*'. And after losing the personal pronouns in the construction, only keeping a general goal-suffix (i.e. a transitivity or passive marker) and introducing markers for agent and goal, this type becomes the Polynesian construction type, with its own possibilities and its own problems of 'passive or active?'.¹⁵

The transitivity-markers bring us to our last point. In Philippine languages we find a neat system of what usually are called focus-types: either the agent, the direct goal, the instrumental goal, the locative goal, or the beneficent goal have the central role in a verbal construction.¹⁶ Usually all the verb-forms, except the actor-focus form, are regarded as passive. But it must be kept in mind only the object-focus form has an active (i.e. actor-focus form) counterpart in which the goal is a direct object. Other languages having suffixes for indicating a relation to an object are different from the Philippine type in two ways. In the first place they show both active and passive forms and constructions with these object-relation markers, whereas the Philippine type languages only have the 'passive' forms. Secondly, they have a less differentiated system, distinguishing only one or two indirect-goal relations. The semantic range of these forms is often very wide. Buli, for instance, has no formal distinction between instrumental and locative in these forms: an '*eat*', ano '*eat with*', topa '*throw*', topo '*throw at*', taping '*to light fire*', tapngo '*shoot at, shoot with*' (Maan 1951:73-79). Balinese has two different suffixes, -in and -ang, but their functions do not seem to be clearly definable: -in usually has locative objects, but also benefactive and even instrumental, whereas -ang marks indirect objects, instrumental, sometimes locative or temporal, e.g. bedbed-ang '*to wrap around*', gebeg-ang '*to rub with*', linggih-in '*to sit on*', ampak-in '*to open for*'.¹⁷ One of the functions of -aka in Wolio is giving the verb in question a subordinated position in the sentences.¹⁸ The prevailing type, however, is that found in languages like Malay, Javanese, Toba, etc., where -i forms are found with a locative goal and (a)kən forms mainly with an instrumental but also with a beneficent goal. In Oceania the difference in use between these two forms have become vague, as both developed in the direction of becoming just transitivity markers without any differentiation in the action-goal relation. The role played by the homomorphy of the locative -i and a former third person object marker -i in the functional reduction of the object-relation marking system can only be guessed.

That also in the Philippine-type languages there is not always a clear-cut set of object-relation categories, might be inferred from the fact that the benefactive (lacking as a separate category in non-Philippine languages) is not fully equivalent to instrumental and locative e.g. in Tagalog.¹⁹ But there is also Malagasy, which shows the Philippine type structure insofar as there is no separate verb-form for active-indirect object, but which in the passive forms only has a difference between direct-goal ('real passive') and indirect-goal forms (traditionally called 'relative forms') with only incidental possibilities of further distinction. The semantic range of the functions indicated by these relative forms is very wide: beneficent, place, time, instrument, price, part of a whole, cause, reason, etc.²⁰

The historical relationship between the Philippine type of verb and the type having active verb-forms with affixes for indirect goals — which take the place of a direct object — might be illustrated by the occurrence of incidental or regular cases of something belonging to one system in the other. In other words: if something normal in the Philippine type were to be found in languages of the other type (and not the reverse) this might be taken as an indication

that the Philippine type was older and that from it the type developed through analogy and rearrangement of morphemes. Active counterparts for an indirect-goal passive is an active verb-form with a prepositional phrase in the Philippine-type languages, or active -i or -(a)kən form in the others. If the latter were to occur – but only incidentally – with a preposition phrase instead of a direct object, this would be what we are looking for.

These forms actually occur: Indonesian *dia menulis surat* 'he writes a letter', *dia menuliskan surat kepada saya* 'he writes a letter to me'. Javanese *klasne digelar* 'the mat is unrolled', but *klasne digelarkan ing jobin* 'the mat is unrolled on the tiles'; *gripe diasah* 'the lead-pencil is sharpened', *kowe diasahake gripe* 'the lead-pencil is sharpened for you', *gripe diasahake ing bata* 'the lead-pencil is sharpened on a stone'. In these cases there seems to be a preference for the -(a)kən forms.²¹ The opposite is also possible: a preposition phrase functioning as focus of an indirect-goal verb in a Philippine-type language: Malagasy.

Tamin'ny alatsinainy no nifanekena tamin-dRakoto
'it was on Monday that a contract was signed with Rakoto'.

These cases, and also the more passive-like use of the transitive-suffix forms in Polynesian, make it very likely that the active -i and -(a)kən forms are a later development and that the special development of the Philippine-type languages mainly consisted in the morphemes, and in the differentiation, of the indirect-goal-focus forms; maybe also in the role of the interplay between verb-forms and definiteness of syntactical elements.²²

NOTES

1. The consequences of these methodological defects are clearly seen in some of R.A. Kern's articles (Kern 1931, 1935).
2. e.g. the restored opposition singular-plural in the pronoun system of Sakaleva and Vezo (Dahl 1951:242).
3. cf. the long and not yet finished discussion on active and passive in AN languages.
4. A good example is Berg (1937).
5. There are too many of them to list them all in the bibliography. The most comprehensive ones are Brandstetter (1912) (now completely outdated), Wolff (1973) and Pawley & Reid (1976).
6. We cite one example of the older publications: Tendeloo (1895).
7. See for Indonesia: Esser (1929) and for Polynesia Krupa (1973), and the literature mentioned by them.
8. See e.g. Berg (1937), where the term 'degrees of activity' is used to cover active, passive, and the in-between forms.

9. To mention just a few: Uma (Esser (1964), Wolio (Anceaux 1952), Biak and the AN languages of Sarera Bay (Anceaux 1961:150-156), Lamaholot (my data from Drs Inyo Yos Fernandes and Drs Aron Mako Mbeté).
10. These lists of names are *not* exhaustive.
11. See e.g. Wolio (Anceaux 1952:28). Philippine languages, and also some languages of Northern Celebes, show comparable forms with some differences in use (Kern 1931).
12. Uhlenbeck's theory was mainly based on American Indian languages (Uhlenbeck 1916 and 1917).
13. Berg (1937), Esser (1964:7).
14. Many are mentioned by Haaksma (1933); other languages are Lamaholot and Wolio.
15. See for the Samoan case: Milner (1962).
16. The vague wording is not without purpose (see Schachter 1976).
17. Kersten (1948:18-28; 1970:54-67).
18. Anceaux (1952:47).
19. Compare, e.g., the treatment of this part of Tagalog grammar by Llamzon (1976:92-106) with that by Ramos (1978).
20. Examples in Malzac (1960:71-74 and 143-150). See also Dahl (1978) and Rajaona (1972).
21. For Dairi Batak I found the following examples of active and passive forms, both having the -(a)kən form in combination with a preposition phrase:

mamberu memereken kepeng baku *'the uncle gives me the money'*,
 kepeng ibereken mamberu baku *'the money is given to me by the uncle'*;
 galuh i iambongken mi teruh (*money that is thrown (+ -ken) to below*)
'that money is thrown down'. (Tinambunan 1980).
22. For this see Bell (1978).

BIBLIOGRAPHY

ANCEAUX, J.C.

- 1952 *The Wolio language: outline of grammatical description and texts.* Verhandelingen Koninklijk Instituut 11. The Hague: Martinus Nijhoff.
- 1961 *The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea.* Verhandelingen Koninklijk Instituut 35. The Hague: Martinus Nijhoff.

BELL, Sarah J.

- 1978 Two differences in definiteness in Cebuano and Tagalog. *Oceanic Linguistics* 17:1-9.

BERG, C.C.

- 1937 Bijdrage tot de Kennis der Javaansche werkwoordsvormen. *Bijdragen tot de Taal-, Land-, en Volkenkunde (BijdrTLV)* 95:1-396.

BRANDSTETTER, Renward

- 1912 *Das Verbum.* Lucerne: Haag.

CLARK, Ross

- 1973 Transitivity and case in Eastern Oceanic languages. *Oceanic Linguistics* 12:559-605.

DAHL, Otto C.

- 1951 *Malgache et Maanyan: une comparaison linguistique.* Avhandlinger utgitt av Instituttet 3. Oslo: Egede-Instituttet.
- 19 *Proto-Austronesian.* Scandinavian Institute of Asian Studies, Monograph Series 15. Lund: Studentlitteratur.
- 1978 The fourth focus. In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 383-393. PL, C-61.

ESSER, S.J.

- 1929 Nogmaals de vervoegde vormen. In: *Feestbundel uitgegeven van het Bataviaasch Genootschap voor Kunsten en Wetenschappen*, Deel I:161-181. Weltevreden: Kolff.
- 1964 *De Uma-taal (West Midden-Celebes): spraakkunstige schets en teksten.* Uitgegeven en van een woordenlijst voorzien door J. Noordwyn. Verhandelingen Koninklijk Instituut 43. The Hague: Martinus Nijhoff.

HAAKSMA, Remy

- 1933 *Inleiding tot de studie der vervoegde vormen in de Indonesische talen.* Leiden: E.J. Brill.

KERN, R.A.

1931 Deutung des -in- Infixes in den Austronesischen Sprachen. *Acta Orientalia* 9:1-58.

1935 De partikel *pa* in de Indonesische talen. *BijdrTLV* 92:5-121.

KERSTEN, J.

1948 *Balische grammatica*. The Hague: W. van Hoeve.

1970 *Tatabahasa Bali*. Ende, Flores: Nusa Indah.

KRUPA, Viktor

1973 *Polynesian languages*. *Janua Linguarum, Series Critica* 11. The Hague-Paris: Mouton.

LLAMZON, Teodoro A.

1976 *Modern Tagalog: a functional-structural description*. *Janua Linguarum, Series Practica* 122. The Hague-Paris: Mouton.

MAAN, G.

1951 *Proeve van een Bulische spraakkunst*. *Verhandelingen Koninklijk Instituut* 10. The Hague: Martinus Nijhoff.

MALZAC, R.P.

1960 *Grammaire malgache*. 4th edn. Paris: Société d'Éditions Géographiques, Maritimes et Coloniales.

MILNER, G.B.

1962 Active, passive or perfective in Samoan: a fresh appraisal of the problem. *Journal of the Polynesian Society* 71/2:151-161.

PAWLEY, Andrew and Lawrence A. REID

1976- The evolution of transitive constructions in Austronesian. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 8/5:51-74.

1979 Also in Paz B. Naylor, ed *Papers from the Second Eastern Conference on Austronesian Languages*, 103-130. Ann Arbor: University of Michigan.

RAJAONA, Simeon

1972 *Structure du malgache*. Fianarantsoa: Ambozontany.

RAMOS, Teresita V.

1971 *Tagalog structures*. Honolulu: University Press of Hawaii.

SCHACHTER, Paul

1976 The subject in Philippine languages: topic, actor, actor-topic, or none of the above? In: Charles N. Li, ed. *Subject and topic*, 491-518. New York: Academic Press.

TENDELOO, H.J.E.

1895 *Maleische verba en nomina verbalia*. Leiden: E.J. Brill.

TINAMBUNAN, T.R.

1980 Akar klausa aktif dar: pasif bahasa Batak Dairi. Unpublished paper for the Seminar on Morphology and Syntax, held at Tugu.

UHLENBECK, C.C.

1916 Het passieve karakter van het verbum transitivum of van het verbum actionis in talen van Noord-Amerika. *Verslagen en Mededeelingen van de Koninklijke Academie van Wetenschappen, Afdeling Letterkunde*, 5^e Reeks, 2^e dl., 187 sqq.

1917 Het identificeerend karakter der possessieve flectie in talen van Noord-Amerika. *Verslagen en Mededeelingen van de Koninklijke Academie van Wetenschappen*, 345 sqq.

WILS, J.

1952 *Het passieve werkwoord in de Indonesische talen*. Verhandelingen Koninklijk Instituut 12. The Hague: Martinus Nijhoff.

WOLFF, John U.

1973 Verbal inflection in Proto-Austronesian. In: Andrew B. Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71-91. *Philippine Journal of Linguistics* Special Monograph 4. Quezon City.

WHERE, O WHERE, HAVE THE LARYNGEALS GONE?
AUSTRONESIAN LARYNGEALS RE-EXAMINED

R. David Zorc

This paper represents a progress report on a monograph (to appear in *Pacific Linguistics*). It is written in the spirit of an Aklanon phrase (mamu:gun kīta 'let's go gleaming in the fields'). I have gleaned from the works of several scholars, and offer here a positive (not negative) criticism of their methodology or approach. From the outset, I must underscore my indebtedness to all of them, particularly at TICAL or in personal correspondence.

1. A BRIEF LOOK IN THE KITCHEN¹

Dyen's now classic monograph (1953) was written in the atmosphere of the discovery of Proto-Indo-European laryngeals. The reconciliation of Dempwolff's somewhat chaotic correspondence sets into PMP *q, *h, and *∅ appeared to justify the title "laryngeals", although it was selected lightly as an explanation of the phenomena observed. In fact, Dyen observed "that either one or both of these proto-phonemes [*q, *h] was phonetically not of a type to which the term 'laryngeals' would normally be applied" (1953:1). Based on currently available reflexes in major Austronesian subgroups, I have suggested – equally lightly – that a title "Some Proto-Austronesian voiceless fricatives" would now be more appropriate, i.e. PMP *q < PAN *[x], PMP *h < PAN *[ç]. Conant (1915:11) had already unified the correspondence of Palau x, Malay h, and Philippine ʔ, but this was either unknown to or ignored by Dempwolff, although cited by Dyen (1953:fn.2). Furthermore, Costenoble (c.1942) foreshadowed Dyen's assignment of Dempwolff's *-h- and *-h as *ʔ, and Dempwolff's *-h- and *-ʔ- as *h; unfortunately this study was only published in 1979 (owing to the efforts and translation of the late Dr Cecilio Lopez).

Although Dyen (1953:1,43,50) did leave considerable latitude for the phonetic interpretation of his *q and *h, it is quite clear that q represents a glottal stop in the citation of data, and that the temporal and geographical separation of Tongan and Tagalog led him to assume this phonetic character for the PMP phoneme. Although Tag and To [ʔ] correspond with Malay [h] < PMP *q, Dyen noted, but left unexplained, the correspondence Tagalic [ʔ], Ml/In [ʔ], and Tongan [∅] (1953:28) which could have led to the reconstruction of PMP *ʔ: *būkaʔ [P136] 'open', *Da:tuʔ [P137] 'chief, ruler', *kakaʔ [P138] 'elder sibling'. [Numbers in brackets refer to data sets in this present study.] Later studies by Dyen

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 111-144. *Pacific Linguistics*, C-75, 1982.

© R. David Zorc

111

(1965a:302f; 1971:36-44) introduced the reconstruction of PAN *ʔ on the basis of a comparison of Philippine and Formosan evidence, but the ML/In evidence was never re-investigated.

Neither *q nor *S were laryngeals (Dahl 1976:35), but do have laryngeal reflexes in many AN languages. The evidence presented thus far for PAN *ʔ has not been convincing (Dahl 1976:37). While data leading to the reconstruction of PAN *H has been limited (1976:130), "it is prudent at our present state of knowledge not to identify all [final *h's] with PAN *S, but as long as correspondences are not found in Formosa, to leave open the question whether they represent *S, some *h, or Ø". (1976:39)

Dyen and McFarland (1970) indicated *ʔ on a large number of PAN reconstructions, particularly in initial and final position (where no other consonant was attested), e.g. *ʔakúʔ 'I', *ʔarúhuʔ *Casuarina*, *ʔasíʔlè 'salt', *bába[ʔh] 'to bring along', etc. Nothofer followed a similar procedure in the reconstruction of Proto-Malayo-Javanic. Since the reconstruction of PAN *ʔ is at issue here, one must be wary of a large number of AN languages that have a phonotactic rule that inserts [ʔ] word or phrase finally (sometimes initially as well), e.g. Sundanese, Kuyonon, Kalamian, Casiguran Dumagat, Itbayaten, Keley-i' Kallahan, Bunun, Paze, Puyuma, Saisiyat, Sediq, Thao, and Atayal.² In this study, considerable weight is attached to the phonemic or the morphophonemic (as opposed to the phonetic or phonotactic) occurrences of laryngeals in various AN languages.³ The importance of morphophonemics cannot be underestimated since inflected forms often preserve phonemes (irregularly) lost in bases. Thus Tsg duwa 'two' (*h > Ø) but ka/whaʔ/an 'twenty' (*h preserved), Aty payat 'four' (*S > Ø) but spat 'eight' (*S preserved), Pai alu 'eight' (*w > Ø) but ka/valu/an 'eighth month, August' (*w preserved).

The choice of 'q' as a symbol for the glottal stop (as opposed to ʔ, ʔ, ʔ) has had unfortunate and confusing consequences for laymen, lexicographers, and linguists, such as the citation of Formosan glottal with ʔ, but Philippine glottal and Malay -k with q within the same article (Dyen 1971:passim).

The confusion of orthographic vs phonemic systems has also led to misinterpretations. Ibans have adopted the Malay orthography even though it does not provide a phonemic fit - it ignores a phonemic final glottal stop and introduces an unnecessary 'o'. A careful and sensitive lexicographer such as Scott (1956) is then forced to give a 'systematic spelling' (i.e. phonemic interpretation) after the traditional spelling. Blust (1970 and articles since) has not cited this final glottal stop, which has herein proven crucial to the reconstruction of PHN and PAN *-ʔ and *-h:

The unexpected appearance of a final glottal stop in some Iban words presents special difficulty. . . Where such a glottal stop is found it is assumed to be of secondary origin, though conditions for its appearance cannot be stated at present. Most of the Philippine languages used extensively in the present comparison also show a final glottal stop in some forms corresponding to a final vowel in Malay, Iban or Javanese. The reconstruction of the 'laryngeals' (*q, *S and *Ø) in all positions is based on a generalization of the criteria presented in Dyen (1953). (1970:110f)

The following sets of data show contrast among four 'laryngeals' in Iban: -Ø (final vowel), -ʔ (final glottal stop), -k (final postvelar unreleased stop), and -h (final voiceless vowel).

mataØ	'eye'	tuaØ	'we-2-inclusive'
mataʔ	'unripe'	tuaʔ	'guardian-spirit'
matak	'pulling' [< batak]	tuak	'palm toddy'
matah	'sloping (land)'	tuah	'lucky'

An extremely large number of minimal pairs appear in Scott, recently corroborated by my work with an Iban informant.⁴ Many of the forms with phonemic glottal stop in final position correspond to cognates among Philippine languages:

Ib aku, Akl ʔăku 'I' > PHN *ăkuØ [P200]
 Ib akuʔ, Akl ʔa:kuʔ 'admit, confess' > PHN *a:kuʔ [P129]

Iban is therefore a criterion language⁵ for the establishment of PHN/PAN *-ʔ.

Nevertheless, following Dyen's conventions for reconstructing *q and dismissing Scott's notation of Iban -q, Blust has ignored this Iban evidence. Although Scott (1956:15) lists Iban badi (badiq) 'a dagger', Blust reconstructs:

76. *badiq 'dagger'. Ceb baríq 'k.o. sickle', Mar badiq 'bolo, machete, knife', Ib badi (*q > Ø unexplained) 'dagger', Ngadha badi 'dagger, sword'. (1973:39 - emphasis mine, see P132)

The plethora of laryngeal correspondences (*S₁, *S₂, *S₃, *x, *X, etc.), certainly never seriously proposed by Dyen as PAN phonemes, has fallen under criticism by several scholars, notably Blust (1978:469f), who raises the point, why has this not been done for PAN *l (with at least three reflexes in Tagalog or for the PAN vowels (allophony of [i]:[e], [u]:[o] has led to phonemic contrasts among several AN languages)? Apparently Dyen's reluctance to characterise the phonetic nature of proto-phonemes or to reduce the inventory to that of a 'real language' is based on the lack of information about subgrouping (e.g. 1953: 50 and 1971:passim). However, it seems that Dyen himself made an appropriate working hypothesis in 1956 - although addressing a different, but related, problem:

The essential point in problems of this type is the basic vocabulary. The correspondence system of a language must be that which applies to its basic vocabulary, taken as a whole; the contrary proposition cannot be tolerated in a genetic comparative treatment. (1956:87)

Granting that many of the most irregular correspondences appear in items meaning 'fire, dog, child, thou, four, tree/wood', it would appear that sorting out these problems would be with recourse to the standard reflexes within each given language's (other) basic vocabulary, rather than with recourse to that language's genetic relationships to other AN languages - except where such information could lead to the identification of loanwords and the direction of borrowing.

2. PURPOSE AND PLAN

In this study, the reconstruction of the PAN laryngeals is based on a systematic progression from two lower-order proto-languages: Proto-Philippine and Proto-Hesperonesian. The comparison of several genetically-remote Ph languages necessitates the reconstruction of four 'laryngeals': PPH *h, *q, *ʔ, and *Ø. These can be aligned with Proto-Indonesian (or Nothofer's PMJ): *h, *ʔ, and *Ø. PHN *h appears to have been lost early in the history of Indonesian languages, and only sporadic evidence can be found in word initial and final position, so that PMJ *h is a reflex systematically, if not exclusively, of PHN

*q. Finally, PAN *q, *S, *Ø, *ʔ, and *H can be reconstructed -- although only the latter two can be considered to have been laryngeal articulations.

In this shortened draft, each phoneme will be dealt with independently, and reconstructions will be labelled according to the time-depth warranted by the currently available evidence. In the absence of *definitive* subgrouping criteria to date, reconstructions are here assigned to putative proto-languages based on the geographical distribution of the witnesses. Hence, only if cognates are found widely distributed in Formosan (North AN), Oceanic (East AN), and Hesperonesian (West AN) are they assigned to PAN. If they are found only in East and West AN, they are assigned to Proto-Malayo-Polynesian (PMP). And, if they are found only in West and North AN, they are assigned to Proto Hesperonesian-Formosan (PHF).⁶ No claim is made that this latter parent language was a reality, as it is possibly equivalent to PAN (based on current subgrouping hypotheses of several scholars).⁷ Rather, this principle is adhered to because certain proto-phonemes are necessitated by Formosan evidence *alone* (*S, *H, amongst others), and in the absence of such evidence no justification can be made, for example, in rewriting PHN or PMP *h as either *S or *H, because the languages used in reconstructing PMP yield a laryngeal inventory identical to that established for PHN (*q, *h, *ʔ, and *Ø).

It must be kept in mind that a reconstruction is founded upon three different kinds of witnesses: *test language* (Pai c < *C, s < *S), *criterion language* (Akl ʔ < *q or *ʔ, Ib -ʔ < *-ʔ or *-h, therefore Akl, Ib ʔ < *ʔ), and *witness language* (helpful in determining the antiquity rather than the phonemic shape of a reconstruction).⁵

3. PROTO-AUSTRONESIAN *q

The reconstruction of PAN *q is straightforward, having been put forward by Dyen (1953) and defended by Dyen (1965, 1971) and Dahl (1976), among others. Healey (1959) has shown that both Kalamian and Tboli (= Tagabili) establish PPH *q in all positions. Historically speaking, the Kalamian dialects⁸ (Karamianenen, Tagbanwa, and Agutaynen) serve as a test language for PPH *q in that [k] is the reflex in all positions, and PPH/PAN *k went to Kl [Ø], although numerous loans from other languages now obscure this phenomenon.⁹ Initially and finally the Tbl reflex is [k], but intervocalically [-h-]. Most other Ph languages reflect *q with [ʔ], although members of the Cordilleran group have final -Ø, and Itb has Ø except in clusters. Nothofer (1975) has shown that PMJ/PIN *h reflects PAN *q, and is established by OJv, Sd, and Ml [h] in all positions; to this may be added the evidence of Ib [h] in final position. At the PMP and PAN level the test languages and respective reflexes are: early Palau data and Bunun (Isbukun dialect) [x], recent Palau data, Tonga, Pazeh, and Kananabun [ʔ], Paiwan, Takituduh Bunun, Ami, Thao, Sediq, and Atayal [q], and Puyuma [h]. Any witnesses with irregular reflexes, due to inflection, innovation, or borrowing, are put into parentheses, but no attempt is made here to subclassify the reconstructed phonemes on this basis.

Table 1: Standard reflexes of some PAN phonemes involving laryngeals

PAN	*q	*ʔ	*h	*s	*s	*∅	*w	*y
Akl	ʔ	ʔ	h	h	s	- ^w y-∅	w	y
Ceb	ʔ	ʔ	h	h	s	- ^w y-∅	w	y
Bik	ʔ	ʔ	h	h	s	- ^w y-h-	w	y
Tag	ʔ	ʔ	h	h	s	-ʔ-h-	w	y
Itb	∅ ʔC	∅ ʔC	h	h	s	∅	w	y
Kal	k	∅	∅	∅	s t	∅	w	y
Tbl	k-h-k	ʔ	ʔ-ʔ-h	ʔ-ʔ-h	s-h-h	-h-	w	y
WBM	ʔ	ʔ	h	h	s	- ^w y-ʔ-	w	y
PPH	*q	*ʔ	*h	*h	*s	*∅	*w	*y
Ib	∅-∅-h	∅-:-ʔ	∅-:-ʔ	∅-:-ʔ	s	∅	∅-w-w	-y-
OJv	h	∅-:-ʔ†	(h)∅-∅	(h)∅-∅	s	∅	w	y-y-
Md	∅	-ʔ-ʔ†	∅	∅	s	∅-h	b-b-	ǰ-ǰ-
ML	h	-∅-ʔ†	(h)∅-∅	(h)∅-∅	s	∅	∅-w-w	-y-
PIN	*h	(*ʔ)	* (h) ∅ (h)	* (h) ∅ (h)	*s	*∅	*w	*y
Pl	x	∅	∅	∅	t	∅	∅-u-∅	-r-
To	ʔ	∅	∅	∅	h	∅	v	∅
PMP	*q	*ʔ	*h	*h	*s	*∅	*w	*y
Aty	q	∅	h†	s	h-h-x	∅-ʔ	w	y
Ami	q	ʔ	h	s	c	∅	w	y
Bun	x	∅	∅	s	s	∅-ʔ	v	ɛ
Tkd	q	ʔ	h	s	s	∅-ʔ	v	ɛ
Kan	ʔ	∅	∅	s	∅	∅	∅	-l-
Pai	q	∅	∅	s	t	∅	v	y
Paz	ʔ	∅	h†	s	z	∅-ʔ	∅	y
Puy	h	∅	∅	∅	s	∅-ʔ	w	y
Sai	ʔ	∅	h†	ʃ	h	∅-ʔ	∅	y
Tha	q	∅	∅	ʃ	t	∅-ʔ	∅	y
Sed	q	∅	x h	s	x	∅-ʔ		y
	[x]	[ʔ]	[h]	[ʃ]	[s]	[∅]	[w]	[y]

† Only in certain environments or on some lexical items.

Where test evidence is not available, criterion languages can assist in establishing certain reconstructions, such as:

- (P01) PHN *qǎbuk¹⁰ 'dust' > Tg ʔǎbok, Ml (h)abok. (NYZ)¹¹
 (P02) PHN *qu:taŋ 'debt' > Akl, Ifg, Ilk, Isg, Sbl, ʔag ʔu:taŋ, (Tɔl ʔuton, Kal utaŋ), Abr, WBM ʔutaŋ, Sd, OJv hutanaŋ, Ml (h)utaŋ. (DNYZ)
 (P03) PHN *tǎqən 'trap, restrain' > Akl, Ceb tǎʔun, WBM, Moŋ taʔan, Ifg tǎʔon, Sd tahən 'trap, snare', OJv tahən 'resist, restrain', Ml tahan 'restrain, set snares; endure, resist', Ib ta:n 'hard-wearing, able to bear, keep, detain'. (BNZ) (See section 6 and note 26.)
 (P04) PHN *tǔquR 'dry' > (Tag tǔyoʔ <), Kpm tǔyuʔ, Abr tuʔug, Sml tohoʔ, Ib tu:r, Ml tohor, Sd tuhur. (CNZ)
 (P05) PHN *lintaq 'leech' > Akl líntaʔ, Btk, Mar lintaʔ, (Kal linta, WBM lintak), Ib, Ml lintah. (ADNSYZ)
 (P06) PHN *lǔdaq 'spit(tle)' > Tag lǔraʔ, Tɔl dulak (M), Ib, Ml mudah. (ABDYZ)
 (P07) PNH *pǎRaŋ 'wring out' > Tag pǐgaʔ, Ceb pǐgaʔ, Ib, Ml pǎrah. (DNY)
 (P08) PMP *Rapuŋ 'rot(ten), brittle' > Ceb gapuʔ, Ml rapoh, Sd rapuh; Fj ravu 'smash up', Ib rapuh 'pile, heap', WBM gapuʔ 'dead wood'. (BNYZ)

In some instances new evidence corrects previous reconstructions:

- (P09) PMP *qaDəp 'front' > Tsg ʔalup/an, Ml (h)adap, Bj ma/hadap, Sd harəp, To ʔao 'front', OJv harəp 'what is placed in front', haDəp 'prevented', To ʔalo 'belly'. (BCNYZ) (Tag hǎrap 'front', Itb harap/ən 'to face', Tsg harap 'foreground', Akl harǎp/un 'near-sighted' < Ml/In.)
 (P10) PMP *qalima:ŋu [crab] > Kal kalimaŋu, Akl ʔalima:ŋu, Pl xemaŋ, Sm alimaŋo, Penchal kimmīŋ. (BYZ)
 (P11) PAN *qalsəm 'sour' > Kal kakləm (A/t→k), Akl ʔáslum (M), Itg ʔalsóm, Ib, (Ml asam), Sd, OJv hasəm, Sm m/asa, Puy ǎrsəm. (DNTYZ)
 (P12) PMP *qa:lun 'wave' > Kal lakun (M), Bik, Ceb ʔa:lun, Ml (h)alon, Ib, NJv, Tb alun, To, Sm ŋ/alu. (BDYZ)
 (P13) PAN *qañud 'flow, drift' > Akl, Ilk ʔa:nud, Kpm ʔanyud, Bj, NgD hañut, Ml (h)añot, (Md añoʔ),¹² To ma/ʔanu, Tkd maŋ/qanuʔ, Ami ma/qalul 'flow, float', Pai qa+ud 'lose'. (ADFTYZ)¹³
 (P14) PMP *qaŋət 'warm' > Ml (h)añat, OJv haŋət, Sd hanət (A/ŋ→t 'warm', To ʔaŋo/aŋo 'dried-out'. (DNYZ) (Tsg haŋat 'uneasiness due to heat'.)
 (P15) PHF *qapəjuø 'gall' > WBM ʔəpəzu,¹⁴ Png ʔapɡú, (Kal apdu, Tɔl hədu/n), Ml (h)ampədu, (h)əmpədu, Jv ampəru, Pai qapədu, Pl xoas. (ADFSTYZ)
 (P16) PAN *qa:puR 'lime' > Akl ʔa:pug, Ifg ʔa:pul, (Kal apug, Tɔl lohoʔ (M), Ml kapor), Pl xaus, Ami qapul 'lime', To n/avu 'put lime'. (BCDSTYZ)
 (P17) PAN *qǎsiN 'salt(y)' > Kal kasin, Tɔl kəhiʔ, Akl, Bon, Isg, Tag ʔǎsin, (Ml, OJv asin), Ib m/asin, Kan ma/ʔáini 'salt', Puy ǎasil 'salty', To m/ahi 'sour, astringent'. (ADNSTYZ)
 (P18) PHN *qǎlət 'interval' > Tɔl kələt 'in succession', Akl ʔǔtut 'gap', WBM ʔələt 'object between two objects', OJv a/hələt 'with an interval of', Sd hələt 'interval', Ml (hə)lat 'alternate'. (BDNSZ)
 (P19) PMP *qi:law 'light; reflect' > Kal kilaw 'intense brightness', Tag ʔi:law 'light', Ml (h)ilaw 'glimmer', Fj ilo/ilo 'anything that reflects: water, glass, mirror'. (BDNYZ)
 (P20) PHN *qipil Intsia [tree] > Kal kipil, Akl, Ilk, Tag ʔi:pil, Bik, Ceb, Han ʔipil, (Ml ipil). (DYZ)
 (P21) PHN *qi(m)pit 'press/squeeze-together' > Kal kipit, Ceb, Ilk, Tag ʔi:pit, Ml (h)impit = (h)əmpit. (BZY) (Note: Akl ʔi<g>pit)
 (P22) PHN *qiRis 'cut/slice-up' > WBM ʔigis 'threaten to stab or spear', Jv ires, Ml (h)iris 'slit, slice, rip open'. (DLZ) (Tag hílis < Ml)

- (P23) PHN *qu:buj 'edible pith' > Kal kuBud, Akl ?u:bud 'palm heart', Ilk ?u:bug 'unspread leaf', (Ml umbut 'palm cabbage'). (BCSYZ)
- (P24) PHF *pǎqiC 'bitter' > Kal pakit, (Tbl he?et), Akl, Bon, Ilk pa?it, Bj, OJv, Sd pahit, Ml pa(h)et, Ib pait, Pl mex/waxed 'bitter', Sai pa?is 'spicy', Kan pa?itsi 'sour' (DNSTYZ)
- (P25) PAN *da:Raq 'blood' > Sbl da:ya?, Ilk da:ra, Ifg da:la, Sml laha?, Ib, Ml darah, OJv rāh, Md Dara, Kan cará?ə, Puy darah 'blood', Pai qaq, Ami lalaaq 'menstrual flow', Pl rasax, Fj ndrā. (ADFNSYZ)
- (P26) PHN *lu:tuq 'cook, prepare (food)' > Kal lutuk, Tag lu:to?, Ilk lu:tu, WBM lutu?, NgD luntoh 'cook', Akl ʔu:tu? 'ripe'. (DSZ)

Numerous examples show Ml/In [h] from PAN *q to have been well retained. Although lost in every other position, it is retained word finally in Iban. Furthermore, although Dyen's original hypothesis regarding PMP *q came under some attack on the basis of his sources, the historicity of Ml orthographic and dialectal [h] is established by substantial external evidence and is a tribute to Dyen's original insights and to Wilkinson's lexicographic method.

There are comparatively few instances of disagreement. Ml fails to reflect [h] in P16, 17, 20, 23, kəntut < *qə(n)tut 'flatulence', ubi 'yam' < *qu:bi(h) (but note Ml humbi), and pusu 'abdomen' < *pu:suq 'heart'. In these few instances the external evidence is weighed more heavily than the lack of an (exact) cognate with [h] in Ml. Conversely, evidence from Ml or other In languages is considered definitive on reconstructions where there are no Kal or Tbl cognates (e.g. P01, P03, P04, P07, P08, P09, P13), or where Kal or Tbl evidence contraindicates the reconstruction of PHN *q (e.g. (P02, P05, P15)). The shift of PAN/PHN *q to PIN/PMJ *h is therefore well substantiated. However, the relationship of PIN/PMJ *h to PHN/PMP *h (and PAN *S or *H) is an entirely different case (see section 6).

4. PROTO-AUSTRONESIAN *S

The reconstruction of PAN (or PHF) *S is only justified by various sibilant reflexes among Formosan languages. The most reliable test languages are Paiwan and Amis with [s]. Saisiyat and Thao, with [ʃ] < *S, also reflect [ʃ] < *C (but under conditions of assimilation rather than of merger with *C).¹⁵ Criterion languages include Bunun, Takituduh, Kananabun, Pazeh, Atayal, and Sediq, all with [s]. These correspondences are related to [h] reflexes in Bisayan (Akl, Ceb, Ogd, Kin, Hil, S-L),¹⁶ Itbayaten,¹⁷ Western Bukidnon Manobo,¹⁸ Bikol (Naga dialect), Hanunoo,¹⁹ Tausug,²⁰ or Tagalog.²¹ Irregular correspondences are put into parentheses and will be discussed at the end of this section; no attempt is made to subclassify the reconstructed phoneme. The following are uncontested reconstructions (i.e. in conformity with the reflexes listed in Table 1):

PAN *Sǎjək 'kiss' (BCTZ) ²²	PAN *SǎpaR 'lay mats' (T)
PHF *Sap(V)Sap 'grope' (TZ)	PHF *Sǎnaw 'wash up' (CTZ)
PHF *Sə(m)pi 'dream' (ADFSYZ) ²²	PHF *Sə(n)Zam 'borrow' (ATZ)
PHF *Sɪnaw 'wash up' (ATYZ)	PAN *Su(w)ab 'yawn' (BCZ)
PAN *i:Səq 'urine' (AFTYZ)	PHF *qami:San 'N wind' (CTZ) ²²
PAN *bǔSək 'hair' (AFTYZ)	PAN *CǎSiq 'sew' (AFTYZ) ²²
PHF *ku:Sa(?) 'go (for)' (TZ)	PHF *lu:Səq 'tear' (TZ)
PHF *ma-buSuk 'drunk' (BZ)	PHF *tǎSəp 'winnow' (ABTZ)
PHF *taSaN(an) 'dwell' (T)	PHF *tu:Sud 'knee(l)' (TZ)
PHF *CiŋaS 'food particles caught between teeth' (TZ)	PHF *RaθuS 'scoop out' (T)
PAN *tǎbuS 'sugarcane' (ABTZ) ²²	PAN *tuqaS 'old (of people)' (ATYZ)

The following reconstructions are either problematic or recently modified:

- (P27) PHF *Sauni 'a while' > Ifg ?awni 'later', Pai nu/sawni 'in a while', ka/sawni, Aty saoni 'a while ago', Ami anu/sauni 'after, later', Tso osni 'right away', Kan saúni 'today', (Paz ?uhuni? 'now'). (TZ)
- (P28) PAN *-Saji(?)²³ 'younger sibling' > Odg ha:li, WBM hazi, (Han ?a:ri?), Md ali?, Ib, Ml adi?, To t/ehi/na, Aty ssue?, Paz sua:zi?, Kuv sua:ni?, Sed suái, Tha ša/šuwá:di?, Ruk ?agí?, Puy wadi?. (ABZ)
- (P29) PHF *SamŭCi *Solanum nigrum* > Itb humti, WBM muti, Bon ?amtí, Isg ?amsí, Pai saməci, RukMg amicu, RukTn amici. (T *-amiCi; Z)
- (P30) PHF *Saŋut 'smell, sniff' > Bik hāŋut, Itb haŋut, Ilk ?a:ŋut 'smell', Pai saŋut 'kiss', (Aty pən/haŋuc 'suck in'). (Z)
- (P31) PAN *Sāpuy 'fire' > Itb, WBM hapuy, Ib, Ml api, To afi (see Table 2 for Formosan data). (ADFNSTYZ)
- (P32) PHF *Sarum 'aroma, smell' > Ml, NgD harum, Pai salum. (A)
- (P33) PHF *Sa(n)təD₂ 'deliver, escort' > Akl hātud, Tag hātí, WBM hatəd, Itb i/hatid (A), Ml (h)antar, Jv atər (Id), Pai saŋəz 'send'. (ADSZ)
- (P34) PHF *Səmay '(cooked) rice' > Akl hūmay, Tsg humay, Ibg, Isg ?əmmáy, Ml imay, Paz sumay, (Ami həmay, Kuv ?əməy). (BCZ)
- (P35) PAN *Səpat 'four' > Tig həp?at, WBM hə?əpat, (Akl ?ap?at, Ceb ?ŭpat, Tag ?a:pat, Itb ?a?pat), Ml əmpat, Tso səptə, RukBd səpátə, Sed səpat, (Sir hpat; for other Formosan data see Table 2), To fa. (ADFNSTYZ)
- (P36) PAN *SəRəC 'tight(en)' > Akl hūgut, Han hu:gut, WBM həgət, Png ?ələt 'tight', Tag hīgit 'pull taut', Itb ma/hyət 'strong', Arosi ro, Tso sərcə 'hug tightly' (note doublet *qəRəC) (BTZ)
- (P37) PHF *Səyup 'blow' > Akl hūyup, Kal əyəp, Tbl m/əyuf, Bon ?ăyup, Tag hi:p/an, Tb iup, NgD tar/iup, (see Table 2). (ADYZ)
- (P38) PHF *Sŷkam 'mat' > Kam hŷkam, Dbw hikam, (WBM ?ikam), Ilk ?ikəm/ən, Tbl ?igam, Kan sikámə, Sar sikamə. (TZ)
- (P39) PAN *Si:paR 'other side; sister-in-law' > Tb ipar, Kan mua/sipáre, Sar i/siparə 'other side'; Akl, Tag, Bik hi:pag, (WBM ?i:pag), Ilk ?i:pag, Ib, Ml ipar, Sa ihe, Fj ra/iva 'sister-in-law'. (ADSYZ)
- (P40) PAN *Si:Rup 'sip, slurp' > Akl hi:gup, WBM higup, Itb h<um>yup, Ml (h)irop, Pai s<əm>iup, (Ami herop); Ulawa iluh/i 'to sup', Nggela ilu 'drink cabbage soup'. (ABDFNYZ) (note doublet *siRəp > Puy sirəp, Ami mi/cirəp).
- (P41) PAN *Su 'thou, thy' > (Tag ?i:yo, Han ka:/wu), Tso na/su, Sed ni/su?, (see Table 2), Kin -u. (ABTYZ)
- (P42) PHF *Su:lij 'sleep with' > Akl, Mas hu:líd, WBM, Tsg hulid, Abr ?ulid, Pai sulid. (Z)
- (P43) PHF *Su:ni [sound of birds] > Akl, Ceb, Tag hu:ni 'chirp', Itb h<um>uni 'to crow', Jv uni, Ami suni 'sound', (Paz huni?) 'chirp'. (TZ) (see Table 2)
- (P44) PAN Su:Nus 'pull out' > Akl hu:nus 'pull out', Tag hūnos 'moult, peel', Ml (h)unus 'draw off, unsheath', Jv unus 'draw out', Ami su+uc, To unuh/i 'pull out'. (ABNYZ)
- (P45) PAN *Sa+ba:Rat 'monsoon' > Akl, Ceb, Tag haba:gat, Itb havayat, (WBM ?əvagat, Han ?aba:gat), Ilk ?aba:gat, Tkd mi/balat, Bun kalim/ba/balat, Sar baratə, Ml, Sd barat 'west', OJv, NJv barat 'west monsoon'. Note: Ami savalat 'E wind' (T). (ADNSYZ)
- (P46) PAN *quSaNap 'scale (of fish)' > Tsg hu?nap, WBM hun?ap, Sai ?öšalap, Pl xolo, To ?uno 'scale', Fu unaf/i 'to scale'. (Z)
- (P47) PHF *qu:Saw 'thirst(y)' > Akl, Ceb, Tag ?u:haw, Kal kuaw, Itb ma/hwaw; Ilk ma/wáw, Pai ma/qusaw, Puy ha/haw. (Z)

- (P48) PAN *Cu:SuR 'pierce' > Akl, Bik, Ceb, Han tu:hug, Itb tuhuy/ən 'to thread needle, skewer, string (beads)', To tu/i 'thread needle', Pai c<m>usu 'to string beads', Bun tusul/un, (Kan c<um>a/cuúru, Aty t<əm>ofo, Sai söhö/ön 'to thread needle'. (TZ)
- (P49) PAN *D₁úSa 'two' > Ceb, Hil duha, Odg ruha, Itb duha, (Tsg, WBM duwa), Ib, Ml dua, OJv rwa, To ua, (see Table 2). (ABCDNSTYZ)
- (P50) PAN *kaSuy 'tree; wood' > Akl, Bik, Ceb ka:huy, Itb kayuh, Ilk ka:yu, WBM kayuh-, Pl gar, To ?a/kau, Fj ka(θ)u, Puy kawí?, (see Table 2), Ib kayu?. (ABCDFNSTYZ), with metathesis, especially yielding PMP *ka:yuh.
- (P51) PAN *LisəSəq 'nit' > WBM lisəha?, Itb lisaha, Kal likəs (M), Tbl kəliah, Ib linsa?, To liha, Fu, Sm lia, Pai ?isəqəs, Bun ?isxus, Tkd ?icqós, Sai i?šiš, RukBd a/lisəəsə. (BCDSTZ)
- (P52) PAN *(Ra)Sina:wa¹³ 'breath(e)' > Akl ginha:wa, WBM gəhinawa, Itb hinawa, Kpm ?ina:wa, To mā/nava, Kan m/uru/ŋisáa (M) 'to breathe', Ib, Ml nāwa 'soul', Tag ginha:wa, Ilk gin?a:wa 'relief, relaxation'. (TZ)
- (P53) PHF *ba:NaS 'man; husband' > Ceb, Hil ba:nah-, Kapuas, Siang bana 'husband', Sai balaš, Paz bałas, Pai vałas 'male (animal)'. (BZ)
- (P54) PHF *ba:ŋaS [tree] > Akl ba:ŋah, Han ba:ŋa *Orania* sp; Kan vaŋásə, Sar vaŋəə, Pai, Ami vaŋas, Sai baŋaš *Melia azedarach*. (TZ)
- (P55) PHF *baRiθuS 'storm-wind' > Akl, Ceb bágyuh-, Jv, Ml bayu, Kl bariw, Sait baŋyoš (see Table 2). (BZ)
- (P56) PAN Călis 'line, string' > Bon, Ilk, Isg tăli, WBM tali, (Tag ta:li? 'tie, rope', Ceb ta:li? 'to tie', Bik ta:li? 'to thread', Ib), Lm, Ml tali, Pai calis, Paz saris, Sai š<in>ă?iš, Puy Tali?, Kan talisi 'rope, cord', Fj tali/a 'to plait' (ABFTYZ)
- (P57) PAN *daqis 'forehead' > Akl dăhi? (M), Ml dahi, Bj dahi?, Ib dai (poetic), (OJv rahi, NJv Dai, To la?e), Pai đaqis, Bun daxis, Tha sá:qis, Aty rqias. (AFYZ)
- (P58) PAN *ku(S)kuS 'fingernail' > Akl kŭkuh, Ceb kŭkuh-, kŭkh-, Itb kukuh, Ml, (Ib kuku), Fj kuku, (Sed kukkux; see Table 2). (ADFYZ)
- (P59) PAN *pa:qaS 'thigh' > Akl pa:?ah, Kal paka, Tbl hahah, (Itb paa), Pl wax, Bun p<in>asax (M), Puy paħa 'thigh', To pa?a/pa?a 'stalk, stem'. (DSYZ)
- (P60) PHF *piRaS 'roe' > Itb piiyah, Bik piġa, Pai bias (with unexplained b, although PAN *b > Pai v). (TZ)
- (P61) PHF *bənSiq 'rice-seed' > Akl, Ceb bínhi?, Bik banhí?, Tag, Han binhí?, Kal binik, Tbl benek, Ilk bin?í, Ib bənih, Ml bəneh, OJv winih, Sai binši?, Bun binsax (M). (DNSZ; T-Fm *binSəq).

With the exception of Puyuma, some Rukai dialects, and Saaroa (*S > Ø) or Siraya (*S > h), the Formosan languages cited here reflect PAN *S reasonably faithfully with a sibilant. Table 2 illustrates the regular reflexes of *S amongst none of these languages. All of the irregularities noted have been put in and are underscored, although only a few of the many non-problematic sets are included. Note that Paiwan alone is without exception; the other languages show an irregular development from *S > h (and some thence to Ø or ?). Since these irregularities are in the basic vocabulary, they have caused considerable concern for some time now (Dyen 1965, 1971, and Tsuchida 1976), and have led to the sub-classification of the reconstruction of *S into as many as nine correspondence sets, some of which have only one member each.

Confronted with this problem, Dahl notes:

In S.At(ayal) s occasionally alternates with h. If the phonemic limit between s and h is vacillating in more languages, the difference is not so great as it seems. (1976:33)

Table 2: Regular and problematic *S cognates amongst Formosan languages; evidence for *S > *H shift

	PAIWAN	BUNUN (Tkd)	AMIS	RUKAI (dl)	KANAKANABU	PAZEH	SAISIYAT	THAO	ATAYAL
*qamiS	qamis	qamisan	kaqamis	-	?amísanə	?amisan	?ämišan	-	qmisan
*CaSiq	cmaqis	mataqáis	mitaqis	cáisi	tumata?ísi	musa?is	šä?iš	šáqiš	smaqis
*Səpi	misəpi	-	-	sipi	-	pisipi?	?išpi?	-	spi?
*Sǎjək	-	sá:k	misanək	-	-	musazək	sazəkən	šmá:ðik	smok
*təbuS	təvus	ci?bus	təvuc	cubúse	təvése	tubus	katbuš	tufu:iš	-
P35	səpaɥ	sasaspát	səpat	spatə	usúpate	səpat	šəpat	špa:t	<u>payat</u> '4' <u>spat</u> '8'
P49	Dusa	dadusa?	tusa	Dusa	cuúsa	dusa?	ruša?	tú:ša?	rusa?
P57	ɟaqis	daqáis	-	-	-	dá·is	rä?iš	šá:qiš	rqias
P27	nusawni	-	anusauni	-	saúni	?uhuni?	-	-	saoni
P41	tisun	?asúh	kisu	kóso	íikasu	?isu?	šu?u	?í:hu?	isu?
P43	-	-	suní	-	umáuni	muhuni?	-	-	-
P34	-	-	<u>həmay</u>	-	-	sumay	-	-	(mami?)
P48	cmusu	matusul	-	-	cumacuúru	-	<u>söhöön</u>	-	<u>lmoho</u>
P152	səti	-	<u>həci?</u>	Sed hi'i Sed xəyi?	?aísi	-	-	-	<u>hi?</u>
P31	sapuy	sapud	-	<u>apúy</u>	<u>apúlu</u>	<u>hapuy</u> <u>hapwi?</u>	<u>hapuy</u>	?á:puy	<u>hpuniq</u> <u>hapuy</u> 'cook'
P58	kaɥuskusan	kuskus	(kinu?əs)	<u>kalukua?</u>	(anuka)	<u>kalikun</u>	<u>kakluköh</u>	<u>kú:ku?</u>	<u>karuh</u>
P50	kasiw	<u>-kavi?</u>	kasuy	<u>kawa?</u>	<u>kaálu</u>	<u>kahuy</u> <u>kahwi?</u>	<u>kähöy</u>	<u>ká:wi?</u>	<u>khuniq</u>
P37	-	<u>ma?iup</u>	<u>mi?iyup</u>	Kuv səmiúp	-	-	<u>hömiup</u>	<u>myú:p</u>	<u>yəmup</u>
P30	səŋuɥ	-	-	-	-	-	-	-	<u>pənhaŋuc</u>
P40	səmiup	-	<u>herop</u>	-	-	-	-	-	-
P55	-	balivus	balíws	-	-	-	ba:yuš	-	<u>bihui</u>

Although this explanation may be adequate for the [h] reflexes in Aty, it is possible to put forward a general historical explanation. It is clear that by PMP times the sibilant reflexes of *S were completely lost outside of Formosa.²⁴ The sound had shifted to PMP *h, and was subsequently lost in Oceanic, and was on the way to being lost amongst Ml/In languages (see 6). It is here proposed that PAN *S was beginning to shift from a sibilant to a laryngeal early in PAN history, and this shift is preserved in certain basic vocabulary amongst northern and central Formosan languages. This shift was arrested in most Formosan languages, but was complete in the ancestor of all PMP languages and at least Siraya of Formosa. (Puyuma, Saaroa, and Rukai may have been similarly affected if the development was *S > *h > *Ø.) It would be more appropriate to set up doublets rather than new proto-phonemes, although one member must be post-PAN but pre-PMP, e.g. PAN *Săpuy > post-PAN *Hăpuy 'fire'.

5. PROTO-AUSTRONESIAN *H

Whereas a partial dialect merger of PAN *S and *H has been proposed, yielding four post-PAN doublets, *Hăpuy (P31), *Hăyup (P37), *ka:Huy (P50), and *kũkuH (P58), it must be stressed that this is not the origin of PAN *H. This phoneme is reconstructed where certain Formosan languages (Tkd, in particular as a test language, or Sai, Ami, Paz, Aty – in descending order of reliability) have an [h] corresponding to [h] among various Ph languages, final [-ʔ] or, less commonly, [-y] in Iban, and (sporadically) initial [h-] in Ml, OJv, or Sd. The presence of cognate forms with [Ø] reflexes in Paiwan (which faithfully shows PAN *S > [s]) makes Pai a special criterion language in the reconstruction of *H (note that Pai [Ø] < *R, *H, *ʔ, *Ø). In the absence of such Pai cognates, similar witnesses are sought among other Formosan languages which would disallow the reconstruction of PAN *S. This phoneme is best attested in final position; forms that generally meet these requirements of Ib, Ph and Fm languages include:

- (P62) PAN *qũmaH 'farm(land)' > Akl, Ceb ʔũmah-, ʔũmh/an, Kal kuma, Ilk ʔũma, Ib umay, Ml (h)uma, Ami, Tkd qumah, Sai ʔõm/õmah, Paz ʔuma/mah, Aty qmayah, Pai quma 'field', Fu uma/ŋa 'dry taro-field', Aty qumah 'to work a field'. (ADSTYZ)
- (P63) PHF *ba:RaH 'live coals' > Itb vayah 'red-hot', Akl, Ceb ba:gah-, Kal bala, Tbl balah, Ib baraʔ, OJv wā, Ami valah, Paz bahah, Aty bagah, Sai bāäh, SaiT baläh, Sed bawah, Sar varaʔa, Kan bara (Tkd, Pai NC). (ADNSTYZ)
- (P64a) PAN *qa+li:maH 'hand' > Akl, Ceb ʔali:mah-, Kal kalima, (Itb lima), Ib limaʔ, Pl xiim, RukMg aríma, Tkd ʔimáh, Pai lima 'hand', Tkd tal/ʔimáh 'to wash hands', Fu, Sm lima 'hand'. (CZ)
- (P64b) PAN *lĩmaH 'five' > Akl lĩmah, (Itb lim, lima), Ib limaʔ, Tkd himáʔ (M), Pai lima, (Ami lima), Fu, Sm lima, Pl w/im. (ADSTYZ)
- (P65) PAN *qani:ŋuH, PMP *qani:nuh 'shadow' > Akl ʔani:nuh, Kal kaninu, Tb halinu, Bima ninu, Tkd qaniŋúh, Bun xaniŋuʔ, (Sai ʔaliŋuʔ). (BCDZ)
- (P66) PHF *qaŋsuH 'stench' > Akl ʔáŋsuh (of wrine), Tkd qancúh (of rotten tubers), Bun ma/xánsuʔ 'stinking'. (Z)
- (P67) PAN *qi:SuH 'shark' > Ođg ʔe:hoh, Agy kiu, Ib iuʔ, Ml (h)iyu, Fj ŋg/io, Pai qisu, (Ami ʔiso). (ACNYZ)
- (P68) PAN *qu:luH 'head' > Akl ʔu:tuH, Kal kulu, Tbl kuluh, To, Fu ʔulu, (Sai taʔõlõh), Pai qulu 'head', Ib uluʔ 'meaning' (Tkd NC; Itb uxu). (ADNSYZ) [Sai l < *l is inexplicable (T).]
- (P69) PAN *baq(ə)RuH¹⁴ 'new' > Akl, Ceb báqʔuh-, Kal baklu, WBM bægu, (Itb vaʔyu, Ib baru), Ml ba(h)ru, Jv wahu, To foʔou, Tkd baqluh, Ami baqluh/ay, Pai vauq/an, Tha faq+uʔ, (?Seđ buurah). (ADTYZ)

- (P70) PAN *ba:RuH 'hibiscus' > Akl ba:guh, (Itb vayu), Ib baru?, Ml baru, Fj vau, To, Fu fau; Pai vau Macaranga tanarius (T). (ADYZ)
- (P71) PHF *băyuh 'to pound (rice)' > Akl, Ceb, Hil băyuh-, Tkd baéúh, Bun báđu?. (Z)
- (P72) PHF *buŋuH 'head' > (Tag bŋŋo? 'skull'), Tkd buŋúh, Ami vunuH, Kan na/vúŋu, Sar vunu?u 'head'. (T)
- (P73) PHF *qasiRaH 'salt' > Chamorro asiga, Pai qatia, Ami cilah, Tkd qacilah, Bun xasila?. (AZ)
- (P74) PHF *Ca:qiH 'faeces' > Akl, Ceb ta:ʔih-, Kal taki, Tbl ke?, Ib tai?, To taʔe, Pl dax, Pai caqi, Ami taiq (M), (Sai säʔiʔ). (ADNSTYZ)
- (P75) PAN *CăquH 'know (how)' > Akl tăʔuh 'don't know', Kal taku, Ml ta(h)u 'know', Ib tau?, Pai caqu 'able', (Itb mananawu 'study, learn'), Tbl hêm/tahuh 'believe', Tkd taquh 'teach', To ma/taʔu 'right hand'. (ADSYZ)
- (P76) PAN *Cali:ŋaH 'ear' > Ceb talíŋh/ug 'listen to', (Itb taliña 'earlobe'), To teliŋa, Tkd taiŋah, Pai caliŋa, (Paz saŋira?, Ami taŋila (M)), teliŋa. (ADSTYZ)
- (P77) PAN *CapaH 'smoke (foods)' > Akl, Ceb ta:pah 'jerk', tăpah 'smoke', Rov tava, Tkd ma/tapha? (M), Pai capa 'smoke', Tso c<m>apo, Kan c<um>a/capa 'to roast over a slow fire'. (ABDTZ)
- (P78) PAN *Cu:maH 'body louse' > Ceb tu:mah-, Itb tumah, Han, Ilk tu:ma, Ib tuma?, To, Sm, Ml tuma, Sai sumäh, (Aty sum/yeq, Sed sum/i?). (ADFNSTYZ) [Paz ʔumah, but ʔ < *C is inexplicable (T).]
- (P79) PAN *CuNuH 'roast on fire' > Jv, Ml, (Ib), To tunu, Ami mu/tuʔuh, Tkd ma/tunuh, Sai sulöh/ön, Pai c<m>uʔu, Kan c<um>a/cunu.
- (P80) PHF *dăkiH 'body dirt' > Akl dăkih, Ib daki?, Ml daki, PMN daki 'id.', (Paz tu/daki? 'dirty', Itb raci? 'rust'). (STZ)
- (P81) PHF *dăpaH 'sole (foot)' > Akl dăpa/dăpah, Ceb lăpa/lăpah-, Kal dapa/rapa, Soc po/lapa, Ib tapa? (A/t → p), Tkd ʔ<al>apáh, Bun d<al>ápa?. (Z)
- (P82) PAN *kitaH 'we (incl)' > Akl, Ceb kítah-, Ib kitay, Ml, To, Fu kita, NJv kitò, Tkd ʔa/táh, (Ami kíta?, Paz, Sai ʔita?, Aty ita?). (ADNSYZ)
- (P83) PHF *nunuh 'breast' > Mlg nunu, (Sbl nu:nu?), Tso nun?u, Sar nuunu?/a, Paz nunuh, Sed nunuh 'breast', Paz mu/nunuh 'suck milk'. (ATZ)
- (P84) PAN *píjaH 'how many?' > Akl, Ceb píläh-, Itb pirah, Tkd piah, Pai pida, To fiha, (Ami pina, Sai piza?). (ADFSTYZ)
- (P85) PAN *si:kuH 'elbow' > Akl, Ceb si:kuh, Itb sicuh, (Ib), Ml siku, Tkd cikuh, (Sai hiku?) 'elbow', To hiku, Fu siku 'end, corner'. (DSZ)
- (P86) PAN *tu:baH [plant: fish-poison] > Akl, Ceb tu:bah-, Ib tubay, NgD tuwe, Fj tuva = nduva, (Sai ta/tuba? (n), t<um>uba? (v) 'id.', Itb tuva/tuva 'drug'). (DNSYZ)
- (P87) PHF *ZaRa:miH 'rice-stalk, straw' > Akl, Ceb daga:mih-, Ilk gara:mi (M), Ib jërami?, Ml jërami, Jv dami, Pai đamia (M). (ADFSZ)
- (P88) PHF *za:waH 'millet' > Akl da:wah, Ib jawa?, NgD jawe, Puy đawa? 'millet', Itb um/rawah '(of grain) to appear out of the ears when nearly ripe'. (ABDZ); Gonda (1973:322) suggests Skt yava- 'barley'.

Tsuchida (1976:132-8) sets down two correspondence sets (*H₁ and *H₂) which agree only in having Akl, Tkd -h. It would appear that Ami, Paz, Sai, and Aty have lost this sound on a number of words, and do not reflect a difference in proto-phonemes (pharyngeal stop and fricative, as Tsuchida suggests). I accept his *ĩnaH 'mother' and *ămaH 'father' (but the *-H here is a vocative ending; *-ʔ, *-y, and *-ŋ also occur in this function, see Blust 1979 and Zorc (1978:94). I have reservations about *qăbu(H) 'ashes', *qaSəlu(H) 'pestle', *bātu(H) 'stone', *ka:li(H) 'dig', and *ku:Cu(H) 'louse' (while Akl and Tkd both reflect *-H, both Itb and Ib reflect *-Ø). I would also require better evidence for his *ĩni(H) 'this', *Libu(H) 'pen', *taRa(H) 'wait', *zaNi(H)

'near', *təlu(H) 'three', *tutu(H) 'pound' (contrast *tuqtuq), *waRi(H) 'sun, day', *wiRi(H) 'left(side)', and my *-aH 'direct passive imperative', *qaba:Ra(H) 'shoulder', *Ca:wa(H) 'laugh', *Du:Ri(H) 'thorn'.

In intervocalic position the following etyma may be reconstructed:

- (P89) PAN *qaluHi:pan 'centipede' > Tag ?alupi:han (M), Ib (l)lipan, Pl xiul, Ml halipan, Sa eluhe, SaiT ?aLuñähipan, Sai ?awñähipan. (Z)
- (P90) PHF *qu:Huŋ [mushroom] > Akl, Ceb ?u:huŋ, Tbl ko?oŋ, guwòŋ, (WBM ?u?uŋ), Ilk ?u?uŋ, Bun xuŋ, Sar u?uŋa, RukMn ?uŋu, Aty qeihoŋ. (Z) (T-*q₄uS₂uŋ)
- (P91) PAN *baHiØ 'woman, female' > WBM bahi, bëhi?an, Akl ba:yih-, Tbl bo?ih, beh, Ami va/vahiy/an, (Itb ma/v<in>ayi), Pai va/vai/an, Puy va/vay/an 'woman', To fe/huhu 'mother (= 'woman of milk'), fe/f<in>e 'woman'. (ABDTYZ)
- (P92) PAN *bu:Hat 'make, work' > Akl, Ceb bu:hat, Kal buat, Ilk bu:?at 'make', Fj vuat/a, Fu, Sm fuat/a 'harvest', (Tha mu/bú:hat 'work') Ib, Ml buat 'load'. (DNZ) [Tha b < *b and h < *H (or *S) is inexplicable (T).]
- (P93) PHF *buHut 'squirrel' > (Tag bu:?ot, bú?ut 'rabbit', Bik bú?ut 'mountain rat'), Ami vuhut, Tkd puhut, Paz buhut, Aty bhut, Sai ka/bhöt, Kan vuútu 'squirrel'. (T)

The cognation of Kan su/kúam/a 'sickness', ma/su/kuámə 'painful' with Tag gúham [skin disease], Ml guam, ruam 'thrush (skin disease)', Kal kuam, Tbl kə/kam Aphten tropica is doubtful, and ambiguously *gu[SH]am (Tsuchida 1976:227 *guS₂am). I accept Tsuchida's (1976:136,181):

- (P94) PAN *lăHud 'sea' > Paz rahat 'west' (= 'seaward'), SaiT Lăhör, Sai ähör 'downhill', hau/ähör 'downstream', Kal laud, Ml laot, (Akl tăwud) 'sea', Fj lau 'Windward Islands', (WBM lawəd 'in midst of', Pai lauD = lauz 'seaward, downhill') as provisional, but exceptionally problematic.

The following may be reconstructed in initial position (see Pl68 also):

- (P95) PHF *Ha:paw 'to top (off)' > Akl ha:paw 'fill to brim', Tag ha:paw 'scum or excess on top', Tkd tun/hapav 'to swim', ma/tiŋ/hapav 'to float up', ?i/hapav 'outside; shallow'. (Z)
- (P96) PHF *HiRaw 'opening, clearing' > WBM higaw/an 'clear-land/trail', Tkd hiláv, Bun ?i?av 'door(way)'. (Z)
- (P97) PHF *Ha:saq 'whet' > Tag ha:sa?, Kal asak, (Han ?a:sa?), Ilk ?a:sa, Mo ?ata?, Ib ansah, Ml asah, OJv a /asah, Tkd ma/hacaq, (Bun ma/sasaq (A/h → s, cf: Bun ?asaz/un), Pai tataq (A/h → t). (ABCDNSYZ)
- (P98) PAN *Ha:liq 'go; move' > Tag hali:- 'come on', Bik ha:li? 'origin; to leave', Ib alih 'turn over', Ml aleh 'move, change (direction)', Kan taku/aí?i 'change direction'. (DNTYZ)
- (P99) PAN *ma/Hataq 'raw, unripe' > Itb ma/hata, Tbl m/atak, Ib m/ata? (M), Ml m/atah, To mata, Kan matá?ə, Pai ma?aq, Sai mantä?, Ilk ?a:ta. (ADNTYZ)

In some cases, evidence is not clear if the reconstruction should be with *S or *H (as in P92):

- (P100) PHF *[SH]u:Ras 'wash' > Akl, Ceb, Tag hu:gas, (Han ?u:gas, Itb uyas/an) 'to wash', Ml (h)uras 'besprinkle', uras 'wet ground', (Sar tara/a/urase 'rinse off'). (DTZ)

Tsuchida (1976:137) reconstructs *H₁ulə[sə] 'garment' on the basis of Tkd huluc, Ami hələt 'garment', Sa ulo 'wrap up', while I find evidence for PMP *qūləs 'woven-garment/blanket' > Kal kukl/un, Ilk ?ūləs, Kpm ?ūlas, Tig ?uləs/ən 'blanket', Ml (h)ulas 'wrapper'. The Formosan forms may be contaminations or

loans of Fm *kuləs/*lukəs (Aty lukus, Tha hú:luθ, Kuv koləs), and should not be directly connected with the Sa form. On the basis of Paz huhas 'vein', Tsuchida reconstructs PAN *H₂uRaC (despite Tkd ʔulat, Sai ka/was, Ami ulat, Akl ʔugat); on the basis of Sai hipih 'cockroach' one might similarly posit PHF *[H]ipəs (despite Tag ʔi:pis, WBM ʔipəs); however, it is prudent to leave the initial consonant undecided for the present. Alongside Akl hūyap, Ifg ʔūyap, Ceb ʔīhap 'to count', there is Sar <um>a/iapə, but this must be ambiguously *[HS]əyap 'count'.

While some of these reconstructions are subject to review and alternative interpretations, the majority of them validate the reconstruction of a PAN (or PHF) *H different from *S, particularly, but not exclusively, in final position. Apart from Formosan evidence, PMP *h must be reconstructed.

6. A NOTE ON THE LOSS OF PMJ/PIN *h

Nothofer maintains: "Dyen's initial *h₁ through *h₄, *q₁ and *q₂, and *S₄ merge in any case in PMJ *h." (1975:167). So many instances of the loss of PHN *h are to be noted that one must posit a chronology of sound shifts: (1) PHN *h was in the process of being lost (i.e. PHN *h > PIN, PMJ *∅) when (2) the shift of PHN *q > PIN, PMJ *h preserved the phonetic character of *[h] on forms that had not yet undergone #1. Furthermore, certain positions appear to have been more 'fragile' than others, so that PIN, PMJ *h (< *h) was lost intervocally and in clusters (I have found no solid evidence for its retention among In/Ml languages in these positions), retained sporadically in initial position (witnessed variously by OJv, Bj/Ml, or Sd), but rather well retained in final position (at least in Iban, where it is reflected as -ʔ; it otherwise appears to have been lost in this position in all other In/Ml languages). The position taken here differs substantially from that of Nothofer,²⁵ and partially from that of Dyen (1953).²⁶

I do not know of any evidence that shows the retention of *h among Ml/In languages in cognates of the following:

PAN *Saba:Rat 'monsoon' [P45] ²⁷	PMP *haDi:ri 'pillar' (DSWZ)
PHN *haRəZan 'ladder' (CDH)	PHN *qana:haw Livistona (DYZ) ²⁷
PAN *qaluHi:pan 'centipede' [P89]	PAN *Sa:jiʔ 'Y-sibling' [28]
PHN *ha:(ŋ)kup 'take with hands' (DZ)	PHF *Sa:(ŋ)kuC 'transport' (BTZ)
PHN *hǎlas [tree]; 'forest' (DZ)	PAN *Ha:liq 'go, move' [P98]
PHN *hǎlin 'transfer' (CZ)	PHN *ha:muk 'attack, rage' (DY)
PHN *ha:ŋus 'breathe hard' (DYZ)	PHN *ha:(m)pin 'pad' (Z)
PAN *Sǎpuy 'fire' [P31]	PMP *ha:(ŋ)saŋ 'gills' (ADSYZ)
PHN *ha:wak 'waist, body' (ADNSYZ)	PHN *ha:wid 'hold, keep' (Z)
PHF *Sǎmay '(cooked) rice' [P34]	PHN *hīkət 'tie up' (DSZ)
PHN *hiluθ 'poison(ous)'	PAN *Si:paR 'other side; sister-in-law' [P39]
PHN *hi:(m)pun 'fry, small fish' (BZ)	PHN *humbak 'wave' (DYZ)
PHN *hu:baq 'change' (DNZ)	PHN *hu:tək 'brain' (DNYZ)
PHF *Su:ni [sound of birds] [P43]	PHN *ba:hu(?) 'smell' (ADHSYZ)
PHN *bǎhal 'ferment(ed)' (Z)	PHN *bu:haŋ 'throw/pour out' (DY)
PHN *bǎhuR 'mix up, mingle' (BNSZ)	PAN *bu:Sək 'hair' (ABTYZ)
PAN *bu:Hat 'make, work' [P92]	PMP *bǎhaq 'flood' (BDYZ)
PMP *Da:hun 'leaf' (DNY)	PHN *guham 'thrush' (DT)
PAN *DūSa 'two' [P49]	PMP *lǎhuk 'mix, blend (food)' (DY)
PHN *ku:hit 'tap; extract' (B)	

PHF *lu:Səq 'tear' (CNTYZ)	PHN *lühəd 'kneel' (YZ)
PHF *ma-buSuk 'drunk' (BCDZ)	PHN *nahik 'go up' (DY)
PHN *ta/həyup 'blow (tube)' (BY)	PHN *timpuhuq 'squat' (BDY)
PHF *tu:Sud 'knee' (DTYZ)	PMP *wa:hiR 'water' (ABCDNYZ)
PMP *qaRu:hu? <i>Casuarina</i> [P130]	PAN *i:Səq 'urine' (AFTYZ)
PMP *pa:huq <i>Mangifera</i> (DNYZ)	PHF *qaS(ə)lu 'pestle' (ACFNTYZ) ¹⁴
PHF *qaŋSit 'stink' (BCTZ)	PMP *bakhaw <i>Rhizophora</i> (DNYZ)
PMP *buRhu? 'jealous' (BDYZ)	PHF *bənSiq 'rice-seed' [P61]
PHN *DahDah 'chest' (DNZ)	PMP *uDəhi 'late, behind' (BCDYZ)

I know of no convincing cognates between Ph and Ml/In languages reflecting [-h-] in both groups. Ceb, Tag tāhan, Bik ta:han 'to bear, endure' have the secondary meaning of Ml tahan < PHN *tāqən [P03] and are loans rather than cognates (Dyen 1953:12 *tahan). Similarly, Tag da:hak, Ilk da:ʔak 'to spit' are loans from Ml dahak, which is from PMJ *Dəhak (Nothofer 1975:72), cf. OJv rəhak, Ib da:k; compare Dyen (1953:12) *[dD]ahak). Under this hypothesis, Tag bŭhoŋ 'deceitful, cunning' must be discarded as a loan from Ml bohōŋ 'lying, untruthful' (Blust 1970:#72). There are a large number of loans among Ph languages from Ml (or other In speech varieties), e.g. at P02, P09, P14, P22, P56. (See also note 26.)

Conversely, I have found only the following examples of the retention of initial *-h-:

- (P101) PMP *ha:ŋin 'wind' > OJv (h)əŋin, Ml əŋen, Akl ha:ŋin 'wind', To əŋi 'blow', Itb hañin 'storm'. (BCDNYZ)
- (P102) PHN *ha:wən 'open space' > OJv hawan 'road, way', Ml awan 'cloud', Akl, Tag ha:wən 'open', Tsg hawan 'to clear', Itb pi/hawan 'interval'. (BDZ)
- (P103) PHN *hi:ləm 'secret, hidden' > Ceb hi:lum, Tag li:him (M); Ml (h)ilam, ilam-2 'dimly visible'. (Z)
- (P104) PHN *hāsək 'cram in; dibble' > Ml (h)asak 'stuff', Ceb hasuk, WBM hasək 'press down (on)', Tag hāsik, Tbl ʔəhək, Ds ŋ/asok 'to dibble'. (BZ)
- (P105) PHN *hu:lur 'fall; drop' > OJv hulur/ulur 'may the fringes be of', Ml (h)ulor 'to pay out (rope), let go', Akl hu:tug 'fall'. (DNYZ)
- (P106) PHN *huyuŋ 'shake, sway' > Ml (h)uyuŋ 'rock, sway', ter/huyuŋ-2 'stagger', Akl hu:yuŋ, WBM huyuŋ 'shake, rattle'. (B)
- (P107) PMP *ha(m)buR 'spread-out/wide' > Ml (h)abur 'to lavish', Ml (h)ambur 'to scatter, litter', Ib abur 'widely known', Tsg hambug 'stout, blooming', ma/habug 'broad (of cloth)', POC *apu 'sprinkle'. (BZ)
- (P108) PHN *halimtaŋ 'crosspiece' > Ml (h)alintaŋ 'crosswise', Akl halimtaŋ, Han halíntaŋ 'step, rung in ladder'. (Z)
- (P109) PMP *hampir 'near(by)' > Ml (h)amper 'near', Ib ampir 'encroach', Tag hampíl 'peak (of heat or coolness)', Ceb hámpil 'pile, to attach', To ofi 'near, close'. (Kuy ampir 'near') (DY)
- (P110) PHN *hantəp 'dense, deep' > Ml (h)antap 'heavy, dense', Akl hántup 'deep within (oneself)'. (Z)
- (P111) PHN *haiq [interjection] > Ml hai 'oh (astonishment or vocative)', Ib aih [interjection: annoyance or incitement], Akl hay? 'well?'. (Z)

Note also PHF *Sa(n)təD₂ (P33), PAN *Si:Rup (P40), PAN *Su:Nus (P44), and PHF *Su:Ras/*Hu:Ras (P100).

There are a number of instances where a final glottal stop (or -y) in Ib reflects PHN *-h (< PAN *-S or *-H), see P50 (M), P51 (M), P62, P63, P64, P66, P67, P70, P73, P74, P75, P78, P80, P81, P82, P86, P87, P88, and:

- (P112) PHN *ǎnuh 'what(ever)?' > Ib anu?, OJv, NJv, Ml anu 'so-and-so, such-and such', Akl ʔǎnuh, Hil ʔǎnuh-, ʔanh-, Itb -aṅuh, akuh (with k for ŋ), Agta ʔənu 'what?', Itb di/nuh 'where?', Itb si/nuh 'who?'. (ADNSYZ)
- (P113) PHF *CīṅaS 'food particles caught between teeth' > Ib tiṅa?, Akl, Ceb tiṅah-, Itb tiñah, Tbl tiṅu?, Pai, Bun ciṅas, Sai šīṅaš. (BCSTZ)
- (P114) PMP *bu:ṅah 'flower; fruit' > Akl, Ceb bu:ṅah-, Tbl buṅuh 'fruit', Ib buṅay, Tb, Ml buṅa, Sm fuṅa 'flower'. (BDSYZ)
- (P115) PMP *būkuh 'joint, node' > Ib buku?, Akl, Ceb būkuh- 'id.', Ml buku 'knot, knuckle', Fj mbuku 'corner, edge'. (DNSYZ)
- (P116) PMP *Dəpah 'fathom' > Ib dəpa?, OJv Dəpa, Md Dəppa(h), Akl, Ceb dūpah-, (Itb maha/dpa), Fu lofa, To ofa. (BDNSYZ)
- (P117) PHN *kənuh [quote particle] > Ib kənu?, Akl, Ceb kūnuh-, (Itb kunu), Ilk kǎnu. (BZ)
- (P118) PHN *ləṅah *Sesamum indicum* > Ib ləṅa?, Ml ləṅa, Tb loṅa, Akl tṅah, Ceb lūṅah-, WBM ləṅa, Tag līṅa. (DZ)
- (P119) PHF *ṅīluH 'set teeth on edge' > Ib ṅīlu?, Ml ṅīlu, ṅīlu, nilu, Akl, Ceb, Hil ṅīluh-, WBM ṅi/ṅīlu, Pai ṅīlu. (DNSZ)¹³
- (P120) PHF *pākūH [fern: *Athyrium esculentum*] > Ib paku?, Ml paku, Akl, Ceb pākūh-, (Tag pāko?), Tbl hoku?, Ami pahko? (M). (DSTZ)
- (P121) PHN *pəñuh [turtle] > Soc ponu, K-C pənu, Tbl hənuh, Ib pəñu?, Ml pəñu, Sml pəno. (DZ)
- (P122) PMP *pa:Rih 'rayfish' > Ib pari?, NJv pé, Ml pari, Akl pa:gih, Ceb, Han, Tag pa:gi, Fj vai, To, Sm fai. (DNSYZ)
- (P123) PMP *saguh [root-crop] sago > Ib sagu? 'balls of cooked sago', Akl saguh, Tag sago, NJv, Ml sagu 'sago, pith', Fj sako/sako 'pudding'. (DNZ)
- (P124) PHN *sǎwah [snake] python > Ib sawa?, Akl sǎwah, Tag sǎwa. (DSY)
- (P125) PHN *sīDah 'they' > Ib sida?, Ceb sīlah-, S-L sīrah-, (Itb sira), NJv sirò. (DNSYZ) (Note Ami caṅra?, Akl sándah (T) < PHF *saṅDah.)
- (P126) PMP *tūlih 'earwax' > Ib tuli? 'having a suppurating ear', Akl tūlih, Ceb ʔa/tulih-, Tag tu/tuli, Fj tule=ndule 'earwax'. (DZ)
- (P127) PHN *zǎzah 'to carry (goods)' > Ib bə/jaja?, Ml jaja, NJv jòjò 'to peddle, carry merchandise', Akl dǎtah, Ceb, Tag dǎlah-, dalh- 'carry, bring', Itb rarah 'load'. (Z)

There are instances where Ib appears to have lost final -ʔ < *-h (P69, P79, P85, P58), but the number and nature of the agreements indicate that Iban is a criterion language for this reconstruction; borrowing, analogy, or independent innovation would not offer a reasonable counter-hypothesis, particularly in the light of the contrast with final -∅ < *-∅ (see section 8).

7. PROTO-AUSTRONESIAN *ʔ

PAN *ʔ is reconstructed on the basis of criterion languages (see Table 1) such as Akl, Ceb, WMB, Tbl, Ib, Md, Jv, or Ml, where it has merged with other phonemes. Formosan languages (with the possible exception of Tkd and Ami) serve only as witness languages because [ʔ] appears to be a phonetic or phonotactic feature of word closure or onset, and has generally merged with PAN *∅. Evidence for PAN *ʔ is clearest and most abundant in final position.

- (P128) PAN *-ǎma? 'father' > Tbl ma?, Moṅ, Abr, Msk, Tsg ʔama?, Sml ʔəmma?, Md əmma?, Ami w/áma?, Pai k/ama, t/ama, To t/ama?/i, Sm t/amā. (ABCDYZ)

- (P129) PHN *a:ku? 'admit, acknowledge' > Akl, Bik, Han ?a:ku? 'accept responsibility', Ib aku? 'admit, confess; treat as', Lm aku? 'take', Ml aku/an 'acknowledgement'. (CNZ)
- (P130) PMP *qaRu:hu? *Casuarina* > Akl, Ceb ?agu:hu?, Ilk ?agu:?u, ?arũ?u, Ib ru?, Ml (h)aru, øru, Tb oru, Fj z/au, Sa s/elu. (BCDYZ)
- (P131) PHF *qila? 'like, note' > Akl ?i:la? 'like, want', Ib ila? 'keep eye on', Sai mya/?ila? 'like, love'. (Z)
- (P132) PHN *badi? [knife] > Ceb bāri?, Mar, Ib badi?, Ngadha badi. (BZ)
- (P133) PHF *b<al>aŋa? [earthenware vessel] > Tag bālāŋa?, Tb balaŋa, Ml bəlaŋa, Ib bəlaŋa? [earthenware vessel], Pai valaŋa, Sir vaŋara (M) 'mortar'. (DYZ)
- (P134) PHN *bapa? 'father' > Lm, NJv, Ml, Sd bapa?, Md əppa? 'father', Ib bapa? 'father-in-law', Han ba:pa? 'uncle', Ilk ba:pa 'parent or parent's siblings', Sbl ba:pa? 'father'. (DZ; A rejects as nursery word)
- (P135) PMP *bi:Ra? 'discharge; semen' > Ceb bi:ga?, Akl biga:~/un, Tsg biga? 'sexual lust', Mar biga? 'semen', Ib bira? 'evacuate, discharge', Sm pia 'smegma'. (BZ)
- (P136) PMP *būka? 'open' > Akl, Hil, Sbl būka?, Tsg buka? 'open', Mar boka? 'untie', Ib, Lm, NJv, Ml, Sd buka?, Md bukka? 'open', Sa hu?e, To, Fu fuke, Sm fu?e 'to uncover'. (DNSYZ)
- (P137) PMP *Da:tu? 'chief' Akl, Hil da:tu?, WBM, Mar, Tir, Tbl datu? 'chief, leader', Ceb da:tu? 'rich; chief', Ib datu? 'nobleman, chief', Ml dato? 'head of family', OJv ratu 'prince', NJv Datu? 'prince, master', Fj ratu 'honorific title (of chiefs)', Sm latu 'captain'. (BDNYZ)
- (P138) PAN *kaka? 'elder sibling' > Tag, Sbl kākā?, Ilk kākā, Han ka:ka?, Ib aka?, RukMg kaka, Ami kaka?, Kuv haha 'elder sibling', Lm, Ml kaka? 'elder sister', Md kaka?, Fj tua/ka 'older brother', Pai kaka 'sibling'. (ABDSYZ)
- (P139) PHF *kəD_i? 'small (amount)' > Akl saŋ/kĩri?, Pai kəDi. (Z)
- (P140) PAN *ki:Ta? 'see' > Akl, Ceb ki:ta?, Tag ki:ta 'see', pa/kita:~/an 'let see', Kpm ?a:kit (M?), Kal ita, Ilk ki:ta; NJv was/kiTò 'clear insight', foresight', To kite 'to appear', Fu fe/kite 'see one another again', Aty m/ita?, Sed k<m>íta?, Sar k<um>a/kita, Paz mu/kitá?, Sai kita?. (ACDNTYZ)
- (P141) PHF *mama? [male kin: MB] > Tag ma:ma? 'MB; sir', Hil ma:ma? 'old man', Ml mama? 'MB', Jv mama? 'father', Lm mama? 'uncle, parent's younger brother', Aty mama? 'uncle'. (BDTYZ)
- (P142) PMP *muda? 'young, unripe' > Tag mu:ra?, Ib muda? 'unripe', Ml muda 'young', To mula 'young (of vegetables)'. (YZ)
- (P143) PHN *naŋka? *Artocarpus* sp. > Bik, Tag naŋká?, Ceb, Tsg náŋka?, Isg, Ilk náŋka, Itb naŋka, Ib, Sd naŋka?, OJv, Ml naŋka. (DNYZ)
- (P144) PHF *Nasi? '(cooked) rice' > Md nasi?, Ib asi?, Ml nasi? '(cooked) rice', Paz mu/taŋi? 'paddy plant'. (Z)
- (P145) PHN *ni:ni? [female kin: address] > Tag ne:ne?, Akl, Ceb ni:ni? (term of address to girls), Ib ini? 'grandmother; (term of address for old women)', NJv nini 'grandmother'. (DZ)
- (P146) PAN *ŋusu? 'mouth' > Bik ŋu:su?, Sar ŋusuu 'mouth', Tag ŋu:so?, Sa ŋudru 'upper lip', Odg ŋu:su? 'jaw'. (DTZ) [Tsuchida 1976:130 *ŋuθuq]
- (P147) PMP *pa:ku? 'nail' > Bik pa:ku?, Tag pa:ko?, Ib paku?, Jv, Ml paku, Fj i/vako. (BDYZ)
- (P148) PMP *pa:lu? 'strike, hit' > Tag pa:lo? 'strike with hand or stick', Ceb pālu? 'hit or strike in punishment', Ib palu? 'hit, beat (with stick, etc.)', Ml palu 'hit hard with a rigid weapon', Tb, Jv palu 'strike', Fj valu, Sa i/pelu 'fight, make war (on)'. (BDNYZ)

- (P149) PAN *puki? 'vulva' > Ceb pu:ki?, Odg, Rom puki?, Tbl ki?, Sml puke?, Ib puki?, NJv, Ml puki 'vulva', Fj mata/vuki 'ulcer on the sole of the foot' = ('eye of the vagina'), Ami (dialect) puki? 'vulva'. (DNTYZ)
- (P150) PHF *sa:ra? 'filter; sift(er)' > Akl sa:ta?, Tag sa:la?, Bik, S-L sa:ra?, Ibg ta:ra, Pai sara. (Z) [For Pai s < *s, see siva 'g' < *siwaØ]
- (P151) PAN *səD₂u? 'hiccup' > Tbl s<ən>du?, Tir sədə? (A), Ib sədu?, Pai təzu, Fj ma/ədedru 'hiccup', Ml sədu 'short broken respiration (of sob or hiccup)'. (BNSZ)
- (P152) PHF *Səsi? 'meat, content(s)' > Kal isi, (Itb a?si), Pai səti, (Ami həci?), Puy isi?, Aty hi?, Sed xəyi?, hi'i 'meat, flesh', Ib isi?, Ml, Jv isi 'contents', Kan ?a/isi 'exist'. (TZ)
- (P153) PMP *ta:bi? [greeting] > Akl, Ceb, Bik ta:bi? (polite particle) 'excuse me, please', Ib tabi?, Ml tabek (greeting), Tag ba:ti? (M) 'to greet', Fj tavi 'salute in welcome, give hospitality to'. (BZ)
- (P154) PHN *ta:ma? 'hit the mark' > Akl, Bik, Ceb, Han, Tag ta:ma?, Kal tama 'correct; hit the mark', Sbl ta:ma? 'correct', Ilk ta:ma 'suitable, timely', Ib tama? 'enter', NJv tò mò 'strike, reach mark'. (DYZ)
- (P155) PHF *təD₂a? 'leftover (food)' > Bik təda?, Oas təra?, Gubat tūda?, Pai təza, (Tag tīra). (Z)
- (P156) PHN *təku? 'bend' > Mar təkə?, Ib təku?. (BZ)
- (P157) PHN *tiku? 'bend' > Akl, Hil tiku?, Mar tiko?, Ib tiku?. (BZ)
- (P158) PMP *wəDa? [existential] > Kin, S-L wəra?, Tag wəla?, WBM waza?, Tbl la?, Ib n/aday, NJv óra 'none, there is not', Ilk ?addá, wəda, Ilt wada, OJv wwara, Ml ada, Md báDá(h) 'there is', Itb wara/wara 'belongings', Fj wara/i 'not be'. (ADNYZ)
- (P159) PHN *zəra? 'warned by experience' > Akl má/dla?, Hil dú/dla?, Kpm, Tag dəla?, Ilk d<ig>rá, Tbl mə/dəla?, hə/dəla?, Ib jəra?, Ml jəra 'warned by experience', Han da:la? 'fear', WMB dəla(?) - 'frighten away'. (BZ)

Intervocalic *-?- is reconstructed between unlike vowels in contrast with *-q- and *-Ø- on the basis of Ph, Ami, and Tkd evidence; hiatus or reduction appear in all other witnesses. Between like vowels the development of a glottal stop is natural and spontaneous (Dahl 1976:37), and hence may have been a phonetic rather than phonemic feature of the proto-language.

- (P160) PPH *ba?ug 'rotten (eggs); sterile' > Han ba:ug, Agy buug 'rotten', Tag bə'og, Han bə'ug 'sterile', Bon ba:ug 'never married', Ilk ba:ug 'ruined (seeds)', Akl ba:ug (v), bə'ug (adj) 'rot (of eggs)'. (Z)
- (P161) PHN *bə(n)ti?is 'calf (leg)' > Kal bisit (M), Tbl tih (aphesis), Msk biti?is, Akl bati:is, Ceb biti:is, Tsg bi?tis, PMN *bəti?is, Ib, Ml bətis, OJv wətis, Md bəttès, NJv wəntés. (CDNSYZ)
- (P162) PPH *bija:u 'winnowing basket' > Akl bila:u, Tag bila:o, Ilk biga:u, Png bigə'u, (Itb bilaw), Kal biraw. (Z)
- (P163) PHF *Ca?as 'high/long' > Akl ta:as, Ceb, Tag tə'as, Tbl taha? (M) 'long, tall'; Ml, Ib atas 'over, above', Md attas 'above, on top', NJv atas 'clearly visible' (M); Puy ?a/Taš 'high, tall', ?i/Taš 'above', Kan ma/cáasə 'long'. (CNZ)
- (P164) PHF *da:an 'old (thing)' > Akl, Han, Il, Png, Sbl da:an, Mar, WBM, Kly da'an, Atta da:n, Isg da/da:n, Itb may/a'dan, Agy daan, Kl da/dan 'old (thing)', PMN *da'an 'old (rice)', Aty ra/ral, Kuvalan za:n 'old (not new)'. (BSTZ)
- (P165) PHN *Da:in 'to jerk (meat/fish)' > Tag da:in 'jerked fish', Kal daiŋ 'dried meat/fish', Ilk da'in 'salted fish dried in sun', Sd de'eŋ 'raw meat cut into slices, spiced, and then dried in the sun', OJv Deŋ,

NJv DènDèn, DèŋDèŋ 'dried meat', Ml daeŋ, (dahiŋ – with secondary h, counter to Sd, OJv, and Kal evidence) 'slicing into strips and drying in the sun'. (BNZ)

- (P166) PPH *ha:ʔun 'remove from fire' > Akl, Ceb ha:ʔun, Tag ʔa:hon (M), WBM haʔun, Ilk ʔa:ʔun, Kal aun. (Z)
- (P167) PHN *harũʔan [mudfish] > Akl haʔũʔan, (Ceb hálwán), Bik harũʔan, Tbl ʔaluʔ, WBM ʔaruʔan, Ml (h)aruan. (Z)
- (P168) PHF *Həʔə 'yes' > Akl hu:ʔu, Kin hə:ʔə/d, Bik ʔu:hu (M), Han huʔ, Msk ʔəʔə, Tsg huʔu/n, NJv həʔə, hə/m 'yes', Pai ʔəʔə 'no' (sem. reversal). (Z)
- (P169) PAN *ka:ʔən 'eat' > Akl, Ceb, Han ka:ʔun, Tag ka:ʔin, Tbl kən, Abr, Btk kaʔən, WBM kaʔan (A), Bik kaʔon, Kal pa/ŋan, Itb k<um>an, Isg ka:n, Ilk ʔag/kǎʔan, mǎ/ŋan, Tkd m/aʔun, Ami k<əm>aʔən, PMN *kan, *ka:n, Ml ma/kan, Sd ha/kan, Ib ma/kay, Fj kan/a, kan/i, To, Fu ka/i. (ABDNSTYZ)
- (P170) PHN *kũʔul [snail] > Akl kũʔuʔ, Hil kuʔul, Ceb ku:ʔul, Sd kuʔul, Md kòʔòl, OJv kũl, NJv kól, (Tag kũhol).²¹ (DNZ)
- (P171) PHN *laʔin 'other, different' > Akl tǎʔin, Ata, Iln, Mam laʔin, Kal lain, Bik, Ceb la:ʔin, Ib, Sd lain, Ml laen, NJv lèn, Md laèn. (DNZ)
- (P172) PHN *laʔun 'long; slow' > Akl tǎʔun, Tag láʔon, Kal luun 'old', Han la:ʔun 'big', Ib laun 'late', Sd laun, OJv a/lon, NJv a/lòn, Ml laon, Md laòn 'slow'. (DNYZ)
- (P173) PHN *lu:ʔək 'bay' > Ceb, Han lu:ʔuk, Ilk lũʔək, lu:ʔuk (A), Moŋ luʔok, Kal læk, Ml tə/luk. (CZ)
- (P174) PHF *paʔən 'bait' > Akl, Ceb pǎʔun, Han pa:ʔun, Tag pa:ʔin, Kal pæn, Itb aʔpan, Mar paʔan (A), Bon pa:/pan, PMN *paʔan, Ib, Ml um/pan, Tb oppan, Tso to/pan/a, Kan pæne, RukMn pene, Pai pan. (DSTYZ)
- (P175) PHN *paʔis 'roast in leaves' > Tag pa:ʔis, Mar, WBM paʔis, Ib pais 'roast in leaves', Ml pais 'spiced fish cooked in a banana leaf', Sd pais 'what has been roasted in hot ashes', NJv pès 'roasted'. (BNZ)
- (P176) PHN *piʔət 'narrow; crowded' > Akl, Bik pi:ʔut, Tag piʔit, Ceb, S-L piʔut, Ceb pí<g>ʔut, Kal piət 'narrow, crowded', Ml mə/miat (< piat) 'to nip and twist (as one tweaks a boy's ear)'. (BZ)
- (P177) PAN *Rabi:ʔih 'night, evening' > Akl gabi:ʔih, Ceb gabʔih-, gabhiʔ-, Tag gab(?)i, Ilk rabiʔi, Ami laviʔi, Aty gbi/an, Sed biy/an, Kal laBii 'night, evening', To efi/afi 'afternoon or early evening'. (DYTZ)
- (P178) PHF *Riʔək 'thresh' > Akl, Ceb, Han gi:ʔuk, Ceb gĩʔuk, Tag gĩʔik, Bik g<ĩn>ik, Tsg giʔik (A), WBM giʔək, Mar gik, (Ml (h)irek), Tir ʔerek, Ilk ʔirik, Png ʔilik, Kan, Sar <um>a/iriki, Tso m/irʔi, Ami mi/ʔrik, RukMg u/iki. (T-*iRik; Z) (metathesis to *ʔiRək is widespread)
- (P179) PPH *ta:ʔəb 'high tide' > Akl, Ceb, Han ta:ʔub, Tag ta:ʔeb, Btk taʔəb, Moŋ taʔab, Kpm, Ilk ʔatab (M), Png ʔatab 'high tide', Kal təəb 'sea'. (CZ)
- (P180) PSP *tĩʔəl > Ceb, Hil, S-L, Iri tĩʔil 'foot', Tbl tiʔol 'bone'. (CZ)
- (P181) PHF *tinəʔun 'weave' > Ami mi/tinuʔun, Tkd tinʔun, Aty, Sed t<əm>inun, Sar t<um>a/tinəənə, RudBd ua/tínunu, Itb tinun/ən, Gad ma/sinun, Tb tonun, Ib, Ml tənun, Mlg ténuna. (NTY)

At least one form appears to have a cluster with *ʔ (i.e. no test or criterion language gives evidence for *q):

- (P182) PMP *bəRʔat 'heavy' > Kal læbat (M), Tbl bəlat, Akl, Ceb búgʔat, Tag bígat, bigʔát, Btk bæʔgat, WBM bəgat, Moŋ bogat, Kpm ba:yat, Sbl bíyat, Ibl bəlʔat, Png bəlat, Ib, Ml, Sd bərat, Tb borat, NgD behat, Md bərraʔ 'heavy', OJv ma/wwat 'become heavy', bwat 'very; weight', Lm biaʔ 'difficult; heavy', Pl xo/brəod, To, Fu, Sm ma/ma/fa 'heavy'. (CDNSYZ)

Tentatively *ʔ clusters may also be reconstructed on the following:

- (P183) PMP *buʔni 'ringworm (herpes)' > Akl, Ceb bunʔi, Han búnʔi, Bik buʔní, Itb vuʔni, Tag bu:ni, Isg būni, Ulawa, Motu huni. (BCDYZ)
- (P184) PMP *kaʔwit 'hook' > Big, Tag, Ilk ka:wit, Akl, Ceb káwʔit (M), Bj, Ib, Ml kait, Ml kaet, Ilk k<all>a:wit, Tb hait, NgD kawit, To kave. (DYZ)
- (P185) PHF *siʔbu 'urine' > Abr, Btk siʔbu, Pal si:bu, Kyp hiʔbu, Bon, Ilk ʔisbú (M), Gad ʔiʔbu, Ifg ʔihbú (M), Sai kǎ/hbuʔ, Tso sifu, Kan íivu. (TZ)
- (P186) PHF *tiʔNaw 'clear (water)' > Akl, Ceb tinʔaw (M), Tag ti:naw, Tha ma/tilaw 'clear (water)', Itb ma/tnaw 'clean'. (TZ)
- (P187) PHN *haʔnəy 'weave, set up warp' > Ceb hanʔay, Han hanʔay, Sd pi/haneʔ, NJv p/ani, Ml məŋ/ani 'to arrange the warp threads on a loom'; Tag ha:nay 'row, line', WBM hanʔəy 'to wind a rope or string in a figure eight around two posts'. (NZ)
- (P188) PHN *sabʔa [plantain] > Kal saBa, Akl, Ceb sábʔa, Bik saʔba (M), Tag, Ilk sǎba, Sbl ha:ʔa, Han sabʔá, Bj, NJv pisaŋ/saba. (CZ)
- (P189) PHN *buʔyŋ 'pot-bellied' > Tag bu:yŋ, Akl búyʔun (A/ŋ → y), Ilk buyʔuŋ, Ml buyuŋ. (BZ)
- (P190) PHN *tuʔmid 'heel' > Tondano, Matanai tuʔmir, Tonsea, Tombulu tuʔmid, Kl tumid, Uma Juman tumir, Ml, NgD tumit, Hv tumutra, (Ib tumbit). (BSZ)

There appear to be only three possibilities leading to the reconstruction of an initial *ʔ-. The first is evidence for the metathesis of an original intervocalic glottal to initial position (e.g. *ʔiRək (P178) or *ʔaCas (P163)). The second is evidence from bound forms with frozen prefixes, where the resulting intervocalic reflexes substantiate an initial glottal after the morpheme boundary (e.g. *si+ʔǎnuh (P112)):

- (P191) PHF *ʔamin '(use) all' > Akl, Kla, Ilk ʔa:min, Bon, Ifg ʔǎmin, Bon ʔamʔin (M?), Isg ʔammín, ma:min, Tkd muʔamín, Ami maʔmin. (Z)
- (P192) PAN *ʔǎnəm 'six' > Akl ʔa/nʔum, Tag ʔa:/nim, Itb a/ʔnəm, WBM hə/ʔənəm, Tkd ʔaʔa/ʔnum, Ami ʔənəm, Fj, To, Fu ono. (ACDNSTYZ)

The third possibility is the evidence of doubled monosyllables which cannot otherwise be established as having begun with *q-, e.g.

- (P193) PHN *[ʔ]ag[ʔ]ag 'sift; sieve' > Tag ʔagʔág = ʔǎgag 'sift(ed)', Ceb ʔ<al>agʔag 'to sift', Kl agag 'sieve (for rice)'. (BZ)

Reluctance to reconstruct this phoneme in initial position is based on the unavailability of morphophonemic evidence for its occurrence outside of the Philippines, note, for example, Sd ʔinum :: ŋ/inum 'drink'.

8. PROTO-AUSTRONESIAN *∅

PAN *∅ is reconstructed in final position on the basis of test languages such as Akl, Ceb, and Ib, and in intervocalic position on the basis of criterion languages. Bs dialects have a morphophonemic final -∅ which is realised in two ways before suffixes: (1) in some dialects (Akl, Hil, Odg) a final high vowel changes to the homorganic semivowel -w- or -y-, while the low vowel is lost and a glottal stop is inserted; (2) in other dialects (Ceb, S-L) a glottal stop is inserted with sporadic instances of final vowel loss. (Zorc 1977:206ff)

- (P194) PAN *qasa:wa \emptyset 'spouse' > Akl, Ceb ?asa:wa 'spouse', paŋ/asáw $\text{?}/un$ 'to marry', Kal katawa, Bik, Bon, Han, Ilk ?asa:wa , Fu ava/ a, Ami ?acawa? . (ACDTYZ)
- (P195) PAN *máCa \emptyset 'eye' > Akl, Bik, Ceb, Han, Ilk, Isg, Sbl máta, Pai maca, Puy maTa ? , Ami mata, Tkd mata ? , To, Ib mata 'eye', Akl, Ceb mat ? - 'wake (x) up', Tag (dial) mata: $\text{?}/an$ 'look around for'. (ADNYZ)
- (P196) PAN *su:su \emptyset 'breast' > Akl, Bon, Ceb, Han, Ilk su:su, To huhu, Pai tutu, Puy susu ? , Ami cucu, Bun susu ? , Sai hōhō ? 'breast', Akl pa/susw-, Ceb pa/sus ? - 'breast-feed (x)'. (ADSTYZ)
- (P197) PHN *Zu:Ru \emptyset 'corner, angle' Akl du:gu 'meet at right angles', ka/rúgw/an, Itb a/ruyw/an, Sbl du:yu, Mar rugu, OJv pa/dū, NJv pa/dō/n, Ml pen/juru, Sd juru ? 'corner, angle'. (DNSZ)
- (P198) PAN *bātu \emptyset 'stone' > Akl, Ceb bātu, Ib batu, Paz batu ? , Itb vatu, Fj vatu 'stone', Akl ka/ba/batw/an 'rocky area', Ceb batū $\text{?}/un$ 'full of rocks'. (ADNSTYZ) [however, compare Tkd batuh, Akl, Ceb bātuh- 'to stone']
- (P199) PAN *ku:Cu \emptyset 'louse' > Ceb ku:tu, Itb, Lm, Ib, Ml, Ami, To kutu, Tha kú:θu ? , Paz, Sai kusu ? 'louse', Akl hi/ηutw/i 'delouse'. (ADNSTYZ) [however, compare Tkd kutuh, Akl ku:tuh, kutu:h/un 'full of lice']

Where such morphophonemic evidence is not available, PAN *- \emptyset may be reconstructed where languages offer evidence for no other phoneme (e.g. P15):

- (P200) PAN *ăku \emptyset 'I' > Akl, Bik, Ceb, Han ?ăku , Bj, Ib, Jv, Ml aku, Ami k/aku, Bun đ/aku? , Aty s/ako, Sm a $\text{?}u$. (ABDSTYZ)
- (P201) PAN *bəli \emptyset 'buy' > Tag bílih, bilh-, Tbl b<əm>lih, Tsg bi:h (all with regular intrusive -h- < * \emptyset), Sml bælli, Ib, Lm, Ml bæli, Sd bæli ? , Sa holi, Fj voli/a, Pai v<n>əli. (DNYZ)
- (P202) PAN *(k)ămi \emptyset > Akl, Ceb, Han kămi, Bj, Ib, Ml kami, Tb hami, Ulawa ami, Ami k/ami, Bun đ/ami? , Aty s/ami. (ABDSYZ)

PAN *- \emptyset - is reconstructed where most AN languages reflect hiatus (with a possible non-phonemic off glide in some languages); although some Ph languages appear to give evidence of *-w- or *-y-, no such evidence is available from any other language group.

- (P203) PAN *i \emptyset a \emptyset 'he/she' > Ceb, Bik, Tag s/īya, Itb s/iya, Odg s/īda, Mam ?iza , Ib, Ml, Tb ia, Ml d/ia, Pl η/i, Jv -e, Fj, To, Fu, Sm ia, Aty h/ia ? , Sai s/ia. (ABDYZ)
- (P204) PMP *ni \emptyset uR 'coconut' > Akl, Bik, Han, Ifg, Itg nīyug, Ilk, Png nyug, Bon ?inyúg (M), Itb niyuy, Odg nīdug, Pl lius, Ib, Ml ñiur,¹³ OJv nyū, Ml nior, Md ñèyòr, ñèyòŋ, Tb niur, NgD ə/ñoh, Fj, Sa, To, Fj niu.
- (P205) PAN *lū \emptyset aq 'expel from mouth' > Tag, Bik luwa ? , Kal ulak (M), Ml iuah 'expel from mouth', Ib kə/luah 'put out', To lua 'vomit', Pai pa/Də/!uaq 'make sound of vomiting'. (ABSYZ)
- (P206) PAN *bu \emptyset aq 'fruit' > Ilk bwa, Blw, Isg būwa 'areca nut', Akl, Ceb bu:wa ? 'spongy growth inside coconut', Jv woh, Bj, Ib, Ml buah, To fua, Puy vuah 'fruit', Kan vuá $\text{?}ə$ 'orange', Pai vuaq [tuber]: (ADNSTY)

The reconstruction of PAN * \emptyset in intervocalic and final position needs no defence, and had indeed been assumed on forms that here have been reconstructed with * ? , *H, or (in a few instances) *S. In the absence of solid morphophonemic evidence for other phonemes (* ? - in particular), it is most prudent at our present state of knowledge to posit it initially as well, and this is done by stating etyma with initial vowels (e.g. P200, P203). Although it is proposed that * ? contrasted with * \emptyset (and *H and *q) in intervocalic and final position, it is possible that it did not occur in initial position or that it did, but

evidence for this may be forever lost to us. Such a contrast would be difficult to distinguish, and therefore I see no advantage in making an assumption that it was clearly *ʔáku over *∅áku, or in marking such ambiguity as *[∅ʔ]áku.

9. A NOTE ON THE PAN SEMICONSONANTS *w AND *y

Dyen (1953:9) recognised the fact that although he treated PMP *w and *y as consonants, "these may turn out to be simply non-syllabic *i and *u", but he generally left the question open, although *y and *w are not reconstructed next to *i and *u respectively. To do this would indicate a prejudice for the distinctive nature of the semivowels for otherwise a writing such as *lawun would imply two occurrences of the phoneme *u." (1953:9) Dahl maintained: "In PAN there was no phonemic difference between *u and *w nor between *i and *y. They were in complementary distribution, *u and *i at syllable summits (syllabics), and *w and *y at syllable limits (non-syllabics)." (1976:17) The result yields a large number of (potential) trisyllabics, even though the majority of AN evidence indicates disyllabics. Often his reliance is on an inaccurate orthography rather than on a careful phonemic representation of data (e.g. Ibg "ualu" = [wǎ́lú] 'eight', Bagobo "uaig" = [wǎʔig], Tir "uajeg" = [wáyeg] 'water', Ilk "uaig" = [wá:ʔig] 'marsh', Kpm "uanan" = [wá:nan], Ibg "jiuanan" = [ziwǎ́nán] 'right(side)'). (Dahl 1976)

In his review of Dahl, Blust (1976b:223) offers some cogent arguments for the preservation of the character of PAN *w and *y as consonants, e.g. loss of final consonants in Amblau (*w and *y included) and the reduction of antepenultimate vowels to [ə] among Bornean languages. Li (1974) has also demonstrated that the reflexes for *y and *w are different from those of *i and *u, strikingly so in some languages.

- (P207) PAN *qa:yam 'domestic animal' > Md ajám 'chicken', Pl xaram 'animal', Bun xadam 'bird', Kan ʔalámə 'meat', Akl ʔa:yam 'dog'. (ACDNTYZ)
- (P208) PAN *ka:yuH 'tree, wood' > Md kaju(h), Fj ka(θ)u, Kan kaáíu, Itb kayuh; Pl gar 'fire; wood'. [See P50; (ACDFNTYZ)]
- (P209) PMP *la:yaR 'sail' > Pl yars, Fj laθa, Odg ya:dag, Sur la:jag, Ml layar.
- (P210) PAN *wǎ́lu∅ 'eight' > Md bá́llu(h), Tkd vauʔ, Pai -Valu-, Ami, To valu.
- (P211) PHF *ka-wǎ́naN 'right(hand)' > Pai kanavaʔ (M), Sai kaʔnal, Paz ʔanan, Ml kanan, Ho havanana, Kan anáne. (ADTY)
- (P212) PAN *ka-wiRi 'left(hand)' > Pai kaviri, Sai kayiʔ, Paz ʔixiʔ, Tkd tana/vilih, Ml kiri, Fj ma/wī. (ADTY)

However, this does not preclude that the origin of some (irregular) w-reflexes is from *u. Reid (1979:18) has suggested an *[ʔ]u topic marker (Ami, Sed, Itb, Ivt ʔu). It is possible that such a marker has become frozen on certain lexical forms, producing otherwise irregular reflexes of *u-Saji(?) 'younger sibling', *u-a:su 'dog', *u-ǎ́nak 'child'; note that the Paz reflex of *w is ∅ (P211-212), yet it has wazuʔ 'dog', sua:ziʔ 'YS'; Tkd v < *w, yet ʔacuʔ 'dog'; Pai v < *w, yet aʔak 'child' (but vatu 'dog'); RukBd v < *w, yet agi 'YS'; Tha ∅ < *w, yet šašuwá:diʔ 'YS'. The role of grammar and morphophonemics in reconstructing PAN is not to be underestimated.²⁸

10. POSTSCRIPT

Part of the attraction of Dyen's monograph (1953:49) was the elimination of several of Dempwolff's doublets. Nevertheless, many doublets must be reconstructed which cannot be resolved. For example, there is evidence for both *SəRəC (P36) and *qəRəC (Ml hərat, Tbl kəlot, Kan ʔérəcə), *hu:tək 'brain' (Bik hu:tuk, Itb hutək) and *qu:tək (To ʔuto), *lawaq 'spider' (Kal lawak, Puy warah) and *lawəʔ (Ib əmpəlawəʔ), *būkaʔ (Pl36) and *būkah (Ceb būkah-, bukh-, Hil būkah-).

Some of these may be resolved on the basis of grammatical derivations. Note, for example, Ib dua 'two', bə/duaʔ 'divide', sə/duay 'you-two'. It would appear attractive to adopt a hypothesis that *bātuØ (Pl98) was a noun stem, while *bātuH was a verb stem, since -h- appears as a morphophoneme in a number of genetically diverse SPH languages (Tag, Ceb, WBM, Tsg, Tbl); but this analysis fails on *ku:CuØ vs *ku:CuH (Pl99), where the exact opposite situation obtains in at least Akl. Blust (1970, 1979) and Zorc (1978:94) have each shown that phenomena such as accent shift and/or *-ʔ, *-H, *-ŋ, *-y served to mark vocatives, which accounts for discrepancies in the final consonants of many kin terms (e.g. *āmaH, *āmaʔ, *āmay, *āmaŋ, *a:maØ 'father'). Accent shift is also noted in commands (Zorc 1978:73) and, parallel to the grammar of vocatives, this may account for the presence of [-h] in the direct passive command form (Akl, Tkd -ah < PHF *-aH).

A greater understanding of the grammar of AN languages is needed in order to determine the role of grammar in sound changes. For example, I have noted that contentives that are otherwise vowel final in Kuyonon and Kalamian end in [-ʔ], but functors (pronouns, demonstratives, numerals, etc.) do not. In S-L and Tsg some functors have undergone the change *s > h, but no similar change is found on a single lexical item. Functors (and this would include kin terms used vocatively) appear to be subject to separate rules, perhaps due to their extremely high frequency, and thus it is not surprising to find *ʔ retained on kin terms in Ml or Jv (lost on other lexemes), or the retention of both *h and *ʔ in Jv həʔə 'yes' (Pl68).

The reconstruction of two laryngeals (PAN *ʔ and *H) is ineluctable, although much more information is needed on the phonemic and morphophonemic status of [ʔ] and [h] in various AN languages, particularly those of Formosa. I encourage and elicit the assistance and assessments of fellow Austronesianists in determining the nature and provenance of laryngeals in PAN.

Shortly after this paper was presented in Bali, Dahl's excellent study (1981) appeared. The reader will note that each of us has addressed certain similar problems (such as the unification of various *q's into one *q, the reconciliation of *S and *H, and the chronology of the shift of *S and *q > [h] in Malayo-Javanic). After several personal communications, Dahl and I have agreed to disagree on the following: one *S (Zorc) versus *S₁ and *S₂ (Dahl), one *H (Zorc) versus *H₁ and *H₂ (Dahl and Tsuchida), and the phonemic status of PAN *ʔ, *w, *y (Dahl is open to discussion, but not yet convinced). I shall address these problems further in my forthcoming monograph.

11. ABBREVIATIONS

A	Assimilation	Kyp	Kayapa Kallahan
Abr	Aborlan Tagbanwa	Lm	Lampung (Way Lima)
Agy	Agutaynen (Kal)	M	Metathesis
Akl	Aklanon (Bs)	Mam	Mamanwa
Ami	Amis	Mar	Maranao
AN	Austronesian	Mas	Masbateño
Ata	Ata Manobo	Md	Madurese
Aty	Atayal (Squiliq)	Ml	Malay(sian)
Bik	Bikol (Naga)	Mlg	Malagasy
Bj	Bandjarese (Ml)	Moŋ	Mongondow
Bkd	Binukid Manobo	Mr	Murik
Blw	Balangaw	Msk	Mansaka
Bon	Bontok	NC	not cognate
Bs	Bisayan	NgD	Ngadju Dayak
Btk	Palawan Batak	NJv	New Javanese
Bun	Bunun (Isbukun)	Ntg	Northern Tagbanwa (Kal)
But	Butuanon (Bs)	Odg	Odionganon (Bs)
Cas	Casiguran Dumagat	OJv	Old Javanese; Kawi
Ceb	Cebuano (Bs)	Pai	Paiwan
Dbw	Dibabawon Manobo	Pal	Palawan
Fj	Fijian	PAN	Proto-Austronesian
Fm	Formosan	Paz	Pazeh
Fu	Futuna	PFM	Proto-Formosan
Gad	Gaddang	Ph	Philippine
Han	Hanunoo	PHF	Proto-Hesperonesian-Formosan
Hil	Hiligaynon (Bs)	PHN	Proto-Hesperonesian (West AN)
Ho	Hova = Malagasy	PIN	Proto-Indonesian
Ib	Iban = Sea Dayak	Pl	Palau
Ibg	Ibanag	PMJ	Proto-Malayo-Javanic
Ibl	Inibaloi	PMN	Proto-Minahasan
Ifg	Ifugao (Batad)	PMP	Proto-Malayo-Polynesian
Ilk	Ilokano	Png	Pangasinan
Iln	Ilianen Manobo	POC	Proto-Oceanic
Ilt	Ilongot (Kakiduge:n)	PPH	Proto-Philippine
In	Indonesian	PPN	Proto-Polynesian
Isg	Isneg	PSF	Proto-South Formosan
Itb	Itbayaten	PSP	Proto-Southern Philippine
Itg	Itneg (Binongan)	Puy	Puyuma
Ivt	Ivatan	Ruk	Rukai
Jv	Javanese	RukBd	Rukai (Budai dialect)
Kal	Kalamian	RukMg	Rukai (Maga dialect)
Kan	Kanakanabu	RukMn	Rukai (Mantauran dialect)
K-C	Kalamansig Manobo	RukTo	Rukai (Tona dialect)
Kin	Kinaray-a (Bs)	RukTn	Rukai (Tanan dialect)
Kl	Bario Kelabit	Sa	Sa'a
Kla	Kalinga	Sai	Saisiyat (Tungho dialect)
Klg	Kalagan	SaiT	Saisiyat (Taai dialect)
Kly	Keley'i' Kallahan	San	Sangir(ese)
Knk	Kankanay	Sar	Saaroa
Kor	Koronadal Bilaan	Sbl	Sambal (Botolan dialect)
Kpm	Kapampangan	Sd	Sundanese
Kuv	Kuvalan	Sed	Sediq
Kuy	Kuyonon	Sin	Sindangan Subanon

Sir	Siraya	Tbl	Tboli, Tagabili
Skt	Sanskrit	Tha	Thao
S-L	Samar-Leyte, Waray (Bs)	Tig	Tigwa Manobo
Sm	Samoan	Tir	Tiruray
Sml	Samal, Sinama	Tkd	Takituduh Bunun
Snl	Sangil	To	Tongan
Soc	Siocon Subanon	Tsg	Tausug
SPh	Southern Philippines	Tso	Tsou
Sur	Surigaonon (Bs)	UAN,	Ur-Austronesian
Tag	Tagalog	Ur }	[Ur=Ger. "Proto"]
Tb	Toba Batak	UJ	Uma Juman
		WBM	Western Bukidnon Manobo

NOTES

1. In 1971 while preparing the "Proto-Philippine finder list" for fieldwork, I accepted many instances of final Ph [-ʔ] as loans or irregularities, but in the interest of being systematic, I marked reconstructions with an underlined zero (0). As my research was in progress, I noted an ever growing corpus of data that agreed in showing a final glottal stop. In 1976 while going through Scott's Iban dictionary, I found a number of agreements with Philippine languages. However, I felt that since Blust had been to Borneo and knew the languages well, his rejection of Scott's phonemics was well founded. By 1978, after continued research into Iban, I felt the agreements amongst Philippine -ʔ and Bisayan -h :: Iban -ʔ were too many and too convincing to be dismissed, especially since Ib and Bs appeared to reflect *-Ø faithfully as -Ø. I revamped all of my data, eliminating the symbol 'q' (except where it genuinely represented a post-velar stop in data or was warranted in reconstructions), re-writing the glottal stop as 'ʔ'; I then began to see the necessity for reconstructing *ʔ as opposed to *q, and began drafting a monograph on AN laryngeals in September 1978. I wish to thank Blust, Dyen, Nother, Sneddon, Harvey, Tchekhoff, Li, Ferrell, and Wolff for their many helpful and positive comments at TICAL, and Charles, Dahl, and Tsuchida for their correspondence concerning this paper.
2. Murik (Blust 1974a) and Uma Juman (Blust 1977c) appear to have gone through a stage where all final *-Ø > -ʔ (sometimes with -n doublets), but synchronically this is no longer the case, so there are vowel-final stems either from *-q or from monophthongalisation:

Mr, UJ aluʔ 'pestle' < PHN *qahluØ
 Mr awaʔ, awan 'spouse' < PAN *qasa:waØ
 Mr, UJ batuʔ 'stone' < PAN *bātu[ØH]
 Mr, UJ bayaʔ 'crocodile' < PMP *buqa:yaØ
 Mr buŋaʔ, (UJ buŋa) 'flower' < PMP *bu:ŋah
 (Mr baa), UJ araʔ 'ember' < PHF *ba:RaH
 Mr bukuʔ < PHN *bũkuh 'joint'

Mr baha, UJ basa 'wet' < PMP *bäsəq
 Mr, UJ ue 'rattan' < PAN *qũøəy
 Mr, UJ laki? 'male' < PMP *la:kiø
 Mr, UJ asa 'whet' < PHF *Ha:saq
 Mr n/uta, UJ uta 'vomit' < PAN *u:taq

A similar phenomenon is observed for Bario Kelabit (Blust 1974d) which shows -h on all forms that came to end in earlier -ø:

Kl təb^huh 'sugarcane' < PAN *təbuS
 Kl mud^hih 'behind' < PMP *uDəhiø
 Kl pəd^huh 'gall' < PHF *qa(m)pəjuø
 Kl barəh 'ember' < PAN *ba:RaH
 Kl matəh 'eye' < PAN *mäCaø
 Kl əb^ha? 'water' < PMP *bāhaq 'flood'
 Kl bada? 'inform' < PHF *bajaq 'know'
 Kl əd^ho 'day' < PAN *qaljaw

3. This is so because one often hears a glottal stop in list intonation which disappears in inflection, e.g.

Kuy dugú? 'blood' :: d<in>uguán 'blood-pudding'
 Kal ?álan? 'buy' :: pa/aləŋ/ən 'make (someone) buy'
 Kal kasáwa? 'spouse' :: kasawá/na 'his wife'
 Itb pasu? 'roast' :: pasu/ən 'roast (it)!'
 Sd ?inum 'drink' :: ŋ/inum 'to drink'

A satisfactory analysis of such phenomena would yield a morphophoneme /ø/ which has a realisation of [ʔ] in specific environments. The glottal stop is not given on data cited from Philippine languages which exhibit this.

4. Challengga Anak Manjan (now Mrs Holt Thompson) from Rumah Tabor, Sungai Sera', Bintulu, Fourth Division, Sarawak. Iban is a fascinating language for the historical linguist in the following areas:

- (1) many instances of semantic reversal (see Blust 1980a),
- (2) -y < *-ŋ, *-n, *-h, *-, as in datay < *dətəŋ 'arrive', makay 'eat' < *ma/ka:ʔən, bukay 'other' < *bukən, panjay 'long' < *panjaŋ, jalay 'road' < *Za:lan, naday 'don't have' < *waDaʔ, umay 'farm' < *qũmaH, etc.
- (3) loss of *C- in functors (including words used vocatively), e.g. udah < *sudah 'already', agi? 'still' < *lagi?, ari 'from' < *dari, ukay 'none' < *bukən, apay 'father' < *bapaʔ, etc.
- (4) vowel length as the result of loss of laryngeals, e.g. ba:h 'flood' < *bāhaq, ba:l 'fermented' < *bāhal, bu:k 'hair' < *bũSək, da:n 'branch' < *daqan, sa:n 'carry on shoulder' < *saqan, ma:r 'expensive' < *māhal, etc.
- (5) final *-R > Ib -ʔ on a few items of basic vocabulary, e.g. iku? 'tail' < *i:kuR, ili? 'downstream' < *qi:liR, kapu? 'lime' < *k+apuR, ai? 'water' < *wa:hiR, təlu? 'egg' < *qi+təluR.
- (6) the retention of *q > h, *ʔ, *H, *S > ʔ in final position only, as taken up in this study.

5. An illustration of criterion, test, and witness language evidence is in order. PAN *Cəbuŋ 'spring (of water)' is reconstructed from Akl tubūr/an, Kal tu/tubd/an, Buih tubūr/an, Ivt a/tbur/an, Paz subud, Pai si/cəvud, Sai ka/sbəz, Fj tuvu, To tufu 'spring', Ilk ag/tũbbóg 'to discharge (water)'. Pai is a test language for *C, *ə, *u, *j in that there is a one-to-one correspondence from each Pai phoneme to a PAN phoneme; Pai is a criterion

language for *b because Pai v < *b or *w. Akl and Ilk serve as test languages for the short penult vowel (with gemination being the manifestation in Ilk, even though *ə has assimilated to the following *u). Paz serves as a criterion language for *C (Paz s < *C, *S) along with Kal, Akl, Buhi, Ivt, Ilk (t < *t, *C, *T); while Sai is a test language for *C (s < *C; *S only if there is a *C in the same form). Fj and To here serve as witness languages and attest to the antiquity of the form, the *ə has assimilated to the following *u. Ilk is also a criterion language for *j (g < *g, *R, *j). Hence, each language contributes something to the reconstruction, and, further, each phoneme of each language has a different status. If we did not have the Pai cognate, PAN *C^ubuj could have been reconstructed (although the addition of Mota tov would have salvaged the *ə), or an ambiguity would have to be indicated, as a PAN *C[^uə]buj (syncope in Ivt and Sai is generally but not always due to the loss of *ə). If we did not have the Oceanic evidence, the reconstruction should be labelled PHF.

6. Tsuchida (1976) was more conservative and labelled such reconstructions as PHN. Dahl (1976) used similar labels, but more to indicate the author of a reconstruction rather than its time depth (UAN = Dempwolff, PMP = Dyen, PAN = Dahl).
7. Dahl (1976:124ff) and Blust (1977b; 1980b) each posit Fm as the highest order AN subgroup from which PMP split. Dyen (1965b:56) did not rule out the possibility of a Formosan homeland. PHN is clearly one subgroup under PMP, but it is not clear whether there are two further offshoots (Central and Eastern) or only one. Tsuchida does not take a firm stand (note 6) because Fm languages generally share a large inventory of forms with Western languages (many of which may be innovations), and only a very small number with Oceanic.
8. I owe a special vote of thanks to Ed Ruch (SIL, Philippines) for sending me all Ntg (Kal) lexical entries he had gathered to date with a [k]. Data on Kal (proper) and Agy were gathered by myself in the field in 1971-72.
9. Note Kal atəl < *kätəl 'itch, ilala < *kila:laə 'know (person)', ulit < *ku:liC 'skin', ian < *Si+ka:ʔən 'fish', bitua < *biCu:ka 'intestines', siit < *sākīt 'sick; pain(ful)', manu < *mānuk 'chicken', bua < b^uSək 'hair'. However, it appears that Kal retains *k between like vowels and in double monosyllables: Kal takaw < *ta:kaw 'steal', taka < PSP *tākaə 'fed up with', kaliŋkiŋ < *k<al>iŋkiŋ 'little finger', kamkam < *kamkam 'get hold of', kəmkəm < *kəmkəm 'hold in fist', kiskis < *kiskis 'scrape off', kuskus < *kuskus 'scrub'. Clearly identifiable loans include: Kal kural 'fence' < Sp corral, tambək 'fat' < Bs tambək, kada 'each' < Sp cada, kalag < CPH kälag 'soul, spirit', kanugun 'too bad' < Bs kanu:gun, etc.
10. Oceanic evidence appears to reflect a doublet (*Răbuk) rather than refute the reconstruction of *qăbuk.
11. Abbreviations refer to the work of the following authors: Z = Zorc

A = Dahl	D = Dempwolff	S = Sneddon
B = Blust	F = Ferrell	T = Tsuchida
C = Charles	N = Nothofer	Y = Dyen

Although such credits are given to the research and data of these authors, I must assume full responsibility for their present shape and meaning. In many instances I have modified the reconstruction of an author quite drastically in the light of evidence for laryngeals, accent (see Zorc 1978), or other phonemes. Numbers preceded by a P- refer to data sets in this paper.

12. Nothofer reconstructs PMJ *qañu| (i.e. *ʔañuj) based on Md aňoʔ (1975:103) and dismisses OJv h<in>añut 'was let afloat', NJv aňot, NgD hañut 'float' as loans from Ml (1975:217, fn.82). Tsuchida reconstructs *q₂añu[zz] (1976:167, 200, fn.109), producing the first instance of a final *-Z. The evidence for *qañuj would be based on Md alone, since NPh languages (which normally reflect *j > g) have d (Bon ʔa:nud, Cas ʔanod, Png ʔǎnur), and together with Pai qaʔud point to PAN *qa:ñud. It is probable that Md aňoʔ is an early loan with the regular development of -t > -ʔ. (Although possibly a typographical error, Nothofer ignores initial Ml (h) and reconstructs *ʔ-.)
13. I do not generally reconstruct *ñ next to *i unless there is strong evidence for doing so. Forms affected include: *niʔa instead of *ña, *qa:niC (not *qa:ñiC), *(Ra)Sina:waʔ (not *ñawa nor *ñiSa:waʔ), *niʔuR (not *ñiʔuR), *ñiʔluh (not *ñiʔluh).
14. WBM does not normally reflect syncope (WBM gəhinawa < *(Ra)Sina:waʔ 'breathe', li səhaʔ < *li səSeq 'nit', həgəzan < *haRəZan 'staircase', ʔəpəzu < *qa(m)pəjuʔ 'gall, bile', pəmənahik < *panahik 'go up') so that some forms are either irregular (perhaps because they involve laryngeals) or reflect genuine consonant clusters in PPH (if not in PHN and PAN: PAN *bǎR(ə)qəŋ 'molar' > WBM baŋəŋ, PAN *Cuq(ə)laN 'condylar bone' (B) > WBM tulan, etc.
15. Li (1978:141) made this observation for Sai; the only case observed for Thao is šáqiš 'sew' < *CaSiq, but observe also the assimilation in Thao *šá:qiš 'face' < *daqiS (Ferrell 1976).
16. For the reconstruction of PBS see Zorc 1977. For purposes of this comparison, all evidence from Bs is taken as a single witness following a similar procedure adopted by Dyen (1953:6f) where he took all Tagalic evidence as one witness. However, Bs evidence is clear and more cogent than Tag.
17. Itbayaten reflects both *S and *H quite faithfully in initial, intervocalic, cluster, and final position, although several instances of h-loss have occurred. This is attributed to influence from Ivt (or other surrounding) h-losing dialects, where Bs, Ib, and Tkd reflect *H or *S (e.g. P69) the reconstruction is considered substantially weak, which is why I am hesitant to accept Tsuchida's *bātuH or *ku:CuH.
18. Western Bukidnon Manobo is generally an excellent witness, except in antepenultimate syllables (ʔəvəgat < *Haba:Rat) or in clusters (ʔandu < *qaS(ə)lu(H)). Data so far assembled appear to indicate a contrast, at least after final a, between *-ʔ (with intrusive -ʔ-) and *-h:

ʔuma 'arrive' :: nə/ʔumah/an 'was met'
 bava 'carry' :: bəvah/a 'carry (it)!'
 dəpa 'fathom' :: dəpah/an 'fathom (it)!'
 duma 'accompany' :: dumah/a 'go with (him)!'
 ʔəsawa 'spouse' :: ʔəsəwaʔ/an 'marry'
 duwa 'two' :: nə/zuwaʔ/an 'two days ago'
 kilala 'know' :: kiləlaʔ/a 'recognised'
 kuwa 'get' :: kuwaʔ/a 'get (it)!'
 ʔupiya 'good' :: ʔupiyaʔ/an 'find (it) good'
 huna 'precede' :: hunaʔ/an 'go ahead of'
 pəmula 'plant' :: pəmulaʔ/an 'planted (in)'

The data are from Elkins (personal communication, 7 May 1980) and are unfortunately not indicated in his dictionary or grammatical sketches.

19. Blust (1974c:138; 154, fn.31) maintained that Han lost *h between unlike vowels, but retained it between like vowels. Rather, Han shows sporadic loss of PPH *h in virtually every environment, and rather fits the theory of lexical diffusion. The 'motivation' for this loss is the surrounding *h-losing languages (Semirara, Kuyonon, Datagnon, and Buid). Note: Han bāha? 'flood', but bā?ag 'loincloth', bāhug 'mix with' but da:wun 'leaf', būhuk 'hair' but ?agu:ʔu Casuarina, ?u:hay 'rice panicle' but dūwa 'two' or ?ūwaw 'thirst', ?aba:gat 'SW monsoon', ?ami:han 'NE wind', ?i:hi? 'urine', ?i:hu 'shark', etc. Thus, the presence of [h] in Han serves as evidence for either *S or *H, but the absence of [h] does not serve as counter-evidence. Note Han binhí? 'seed', bakhaw [tree], ginha:wa?.
20. Tausug generally reflects PAN *S and *H faithfully except in final position where *-∅, *-H, and *-S merge in Tsg -h. Tsg is one of the few non-Fm languages that reflects *Si- 'instrumental passive' as hi- (rather than ?i-), and although *S is lost in Tsg duwa 'two', it is retained in ka/wha?/an 'twenty' (with regular loss of *D).
21. Tag is perhaps the poorest witness for PPH *h and PAN *S, *H because of numerous secondary developments and loans. Like Tsg, Tag has merged *-∅, *-H, and *-S, so that all but a handful of vowel-final forms are inflected with -h- before suffixes. Further, PAN *l has shifted to Tag [h] on some forms, e.g. ba:hay < *bālay 'house', sa:hiŋ 'resin' < *sa:ləŋ, sāhig 'floor' < *saləR, sa:hod 'catch drippings' < *sa:lud, bu:ho? [thin bamboo] < *bu:lug, bi:his 'change' < *bələs, si:haŋ 'chin' < *sələŋ, ?u:hod 'worm' < *qu:ləj, etc. It is posited that this development was through a shift *l > ∅, with -h- insertion on analogy with morphophonemic final -h (Dyen 1973). A similar development is posited for Tag kūhol 'snail' < *kūʔul 'snail' < (P170).
22. See Table 2 for Formosan data leading to this reconstruction.
23. A PAN *u topic marker is posited as frozen on this and other forms with otherwise irregular evidence for *w (see end of section 9). Blust (1979:233) reconstructs *Sua(n)ji and Dahl (1976:80) *[qu]a(ŋ)ji. Whereas Formosan languages appear to have metathesised *S to final position on some forms (*būSək > Fm *bukəS 'hair', *CāSiq > Fm *CaqiS 'sew', *tāSəp > Fm *tapəS 'winnow', *lisəSəq > Fm *lisəqəS 'nit'), they possibly metathesised an original *u+Saji? 'outwards' into initial position, yielding Fm *Suaji. This hypothesis accounts for the subsequent appearance of forms like PHN *hāji? (from the proposed *Saji(?)) and PHN *waji? (from *u+Saji(?) with *S loss).
24. Charles (1974:490ff) and Dahl (1976:125, 130) have each noted difficulties with Blust's vowel deletion hypothesis whereby a proposed PAN *S came to stand in a cluster with some consonant, yielding reflexes such as implosive stops (Bintulu), aspirated voiced stops (Kelabit), or s (Kiput). Dahl shows that only *buSək and *təbuS (but not *təbuSu) conform with Formosan evidence, but eight reconstructions (*qa(m)pəju∅, *bəjbəj, *dakdak, *lūdaq, *ba:baw, *buRuk, *qa(l)jaw, *D₂əmD₂əm) "are without traces of PAN *S". Three forms can only be justified with *h (not *S) on available evidence: PMP *bāhaq 'flood', PMP ūDāhi∅ 'late, behind', and PHN *bāhaR 'loincloth'; one must be reconstructed with *q (PHN *dāqən 'press' > Tag dīʔin, S-L, Bik dūʔun, Tbl dəkən, Kl də/dhən); and one should probably be reconstructed with *q (PHN *Dəqak 'spit' > PMJ *Dəhak, see section 6 herein). Most cases of these strengthened grade reflexes occur after *ə, and all occur after a historically reconstructable short penult vowel (*qījuŋ, *bīlaq, *lūdaq). Such

short vowels led to *sporadic* gemination in Madurese (Zorc 1978:97f), and probably to such disparate reflexes in the North-Sarawak languages. Although considerable support for the hypothesis was gained by apparent *S loss in *five forms* [*buSuk (which is from *bŪSək with assimilation of *ə → *u), *bāhaq (no justification for *S), *dahək (which may have been *Dəqak, and which has no justification for *S), *bāhaR (no justification for *S), *ŪDəhi (no justification for *S) (Blust 1974), surely the Fm counterevidence on *eight forms*, and the need to reconstruct *q or only *h in several other instances does not make the hypothesis compelling. The appearance of doublets in Sarangani Manobo, baha? and ba?ba? 'mouth' need not be evidence for *S (or even *h) since (a) SarMb is basically an *h-losing language, (b) ba?ba? is the regular Manobo development from *baqbaq, and (c) baha? may be an early loan from Bilic languages where the -h- is a reflex of *b, i.e. *baqbaq > Bilic *baba? > *baha? > SarBlaan ba? 'mouth'.

25. Nothofer (1975:166-168) relates PMJ phonemes to PAN as follows:

PMJ *h- < PAN *h-, *q-, *S₄-
 PMJ *-h- < PAN *q-, *-Q₁-
 PMJ *-h < PAN *-h, *-q, *-Q₂

The only instance of final *-h is Dyen's *rapuh (PMJ *rapuh) 'brittle' which should be PMP *Rāpuq (P08). Looking at these correspondences from 'the top down' we get (according to Nothofer's reflexes with *rapuh corrected):

PAN *h, *S > PMJ *h- -Ø- -?
 PAN *q, *Q > PMJ *h- -h- -h
 PAN *?, *x > PMJ *?- -Ø- -?

It is clear that only PAN, PHN *q is retained systematically, and PAN *H and *S > PHN *h has been subject to some diverse changes as explained herein (section 6). The status of PMJ *? on the basis of Sd ? is also open to question, but may be maintained on the evidence of Md, Ml, Jv, Lm, and Ib. Nothofer (1975:24) mentions that "in West Madurese final h occurs in contrast with final vowel, but only before pause. In the citation of Md material this contrast is indicated by writing a final h in parentheses, e.g. buđi(h) 'back'". Historically this appears to be a prosodic feature, much like the development of final Sd -?. Forms reconstructed with *-Ø have this -h: Md mata(h) < PAN *māCaØ 'eye'. Therefore no direct evidence for PMJ, PIN, or PHN *-h is given by West Md.

26. Dyen posited the retention of PMP *-h- only between like vowels (1953:12), citing Ml dahak 'phlegm' (M of *Dəqak) and tahan 'restrain, endure' (*tāqən). He took Ml bah 'flood' (< *bāhaq) as an exception. He recognised the total loss of PMP *-h (17f), and the sporadic loss of PMP *h- in some Ml cognates (32f). Unfortunately, many of the Tag forms cited with h- can be shown to be loans from Ml, and the correct reconstruction is with *q (Tag hala:man < *qala:men, Tag hi:la < *qila, Tag hi:līs < *qīRīs, hali:maw < *qa+Rim?aw, Tag hārap < *qaDəp, ha:tol < *qatu[Rr], etc.).
27. Blust (1978:26f) posits the development of PHN *habe:Rat > Ml barat as regular, in that *a₃ > ə and *h (< *q or *S/*H) was lost if it did not precede a prenasalised stop (thus Ml təlur < *qīCəlur 'egg', but Ml (h)əmpədu < *qa(m)pəjuØ 'gall, gall-bladder'). This change (loss of *ha₃-) preceded the change *-aba- > Ml -awa-. I agree with Blust in this observation, but I have noted that *qa₃- was not always lost; despite *qanibuŋ > Ml niboŋ *Oncosperma*, there is Ml halipan < *qaluHi:pan (P89), Ml halia < pre-Ml

*qaluya < *lǎquya 'ginger', Ml halaman < *qala:mən 'lawn, grass, weeds', Ml hanau < *qana:haw 'sugar palm', Ml (h)aru < *qaru:hu? (Pl30). It would thus appear that Ml generally tended to retain h < *q, but not h < *S/*H.

28. Note further that Ami (which has a productive ʔu marker) has forms such as wáma 'father', wina 'mother', wikul 'tail', along with wacu 'dog'. To date no PAN *w- has been proposed on the reconstructions *ama, *ina, or *i:kuR, and need not be if these forms derive from *u+ama, *u+ina, *u+i:kuR. James Collins' fine paper at TICAL (Prothesis in the languages of Central Maluku: an argument from PAN grammar, see pp.187-200 in this volume) surveys precisely this kind of area. Therein he proposes that PAN *si and *u markers (and possibly an *i marker as well) have led to otherwise irregular reflexes in the languages he studies. The fact that we cannot yet precisely ascertain the function and meaning of these markers does not obviate the need for reconstructing them, as substantial and widespread evidence indicates the existence of such markers in genetically diverse Austronesian languages today. Nor is the appeal to such morphophonemic or grammatical evidence *ad hoc* – at least no more *ad hoc* than the reconstruction of a new PAN phoneme (*W). In each case scholars are attributing an anomaly observed amongst daughter languages to a system of the proto-language, either its phonology or its grammar.

BIBLIOGRAPHY

BLUST, Robert A.

- 1970 Proto-Austronesian addenda. *Oceanic Linguistics (OL)* 9/2:104-162.
- 1972a Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages: a preliminary list. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 4/1:1-43.
- 1972b Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages'. *WPLUH* 4/8:1-17.
- 1973 Additions to 'Proto-Oceanic addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic languages' - 2. *WPLUH* 5/3:33-61.
- 1974a A Murik vocabulary. *The Sarawak Museum Journal* 22:153-189.
- 1974b Proto-Austronesian syntax: the first step. *OL* 13:1-5.
- 1974c The Proto-Austronesian word for "two": a second look. *OL* 13:123-161.
- 1974d The Proto-North Sarawak vowel deletion hypothesis. Ph.D. dissertation, University of Hawaii. [Precis distributed at FICCAL: *WPLUH* 5:91-149.]
- 1976a Dempwolff's reduplicated monosyllables. *OL* 15:107-130.
- 1976b Review of O.C. Dahl *Proto-Austronesian*. *Language* 52:221-237.

- 1977a Austronesian culture history: some linguistic inferences and their relations to the archaeological record. In: Soepomo Poedjosoedarmo, ed. *Miscellaneous studies in Indonesian and languages in Indonesia*, part 3:25-37. NUSA 4. Jakarta: Badan Penyelenggara Seri NUSA.
- 1977b The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *WPLUH* 9/2:1-15.
- 1977c Sketches of the morphology and phonology of Bornean languages 1: Uma Juman (Kayan). *Papers in Borneo and Western Austronesian linguistics* 2. *PL*, A-33:9-122.
- 1978 Review of Thomas A. Sebeok, ed. *Current trends in linguistics*, vol.8: *Linguistics in Oceania*. *Language* 54:467-480.
- 1979 Proto-Western Malayo-Polynesian vocatives. *Bijdragen tot der Taal-, Land- en Volkenkunde (BijdrTLV)* 135:205-251.
- 1980a Iban antonymy: a case from diachrony? In: D.J. van Alkemade et al, eds *Linguistic studies offered to Berthe Siertsema*. Amsterdam: Rodopi.
- 1980b Early Austronesian social organization: the evidence of language. *Current Anthropology* 21/2:205-247.
- 1980c Notes on Proto-Malayo-Polynesian phratry dualism. *BijdrTLV* 136:215-247.
- CHEN, Matthew Y. and William S.-Y. WANG
- 1975 Sound change: actuation and implementation. *Language* 51/2:255-281.
- CONANT, Carlos E.
- 1915 Notes on the phonology of the Palau language. *Journal of the American Oriental Society* 35:1-15.
- COSTENOBLE, Hermann
- 1979 *Dictionary of Proto-Philippine*. Translated by Cecilio Lopez, edited by Ernesto Constantino. Diliman, Quezon City: Archives of Philippine Languages and Dialects.
- DAHL, Otto Christian
- 1976 *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series 15. London: Curzon Press. (1st edn 1973, Lund.)
- 1981 *Early phonetic and phonemic changes in Austronesian*. Oslo: Universitetsforlaget.
- DEMPWOLFF, Otto
- 1938 *Vergleichende Lautlehre des austronesischen Wortschatzes*, 3: *Austronesisches Wörterverzeichnis*. *Zeitschrift für Eingeborenen-Sprachen* 19. Hamburg: Friederichsen, De Gruyter.
- DYEN, Isidore
- 1953 *The Proto-Malayo-Polynesian laryngeals*. Baltimore: The Linguistic Society of America.
- 1956 The Ngaju-Dayak "old speech stratum". *Language* 32/1:83-87.
- 1965a Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285-305.
- 1965b A lexicostatistical classification of the Austronesian languages. *International Journal of American Linguistics* Memoir 19. Baltimore: Waverley Press.

- 1971 The Austronesian languages and Proto-Austronesian. In: Thomas A. Sebeok, ed. *Current trends in linguistics*, vol.8:5-54.
- 1973 Tagalog reflexes of Proto-Austronesian *l. In: A.B. Gonzalez, ed. *Parangal kay Cecilio Lopez*, 3-7. Manila: Linguistic Society of the Philippines.
- DYEN, Isidore and Curtis D. McFARLAND
- 1970 Proto-Austronesian etyma constituting an Austronesian cognate finder list. Yale. Mimeographed.
- EGEROD, Søren
- 1965 An English-Atayal vocabulary. *Acta Orientalia* 29:203-220.
- FERRELL, Raleigh
- 1968 The Pazeh-Kahabu language. *Bulletin of the Department of Anthropology and Archaeology, National Taiwan University*, 31-32:73-97.
- 1969 *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Monograph 17. Taipei: Institute of Ethnology, Academia Sinica.
- 1979 Phonological subgrouping of Formosan languages. In: Paz Buenaventura Naylor, ed. *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*, 241-254. Ann Arbor: University of Michigan.
- 1982 *Paiwan dictionary*. PL, C-73.
- HEALEY, Alan
- 1961 Dyen's laryngeals in some Philippine languages. *SILUND (Summer Institute of Linguistics, University of North Dakota)* 5:53-64.
- HO, Dah-an
- 1978 A preliminary comparative study of five Paiwan dialects. *Bulletin of the Institute of History and Philology (BIHP)* 49/4:565-681. [English abstract p.681.]
- LI, Paul Jen-kuei
- 1974 Alternations between semiconsonants and fricatives or liquids. *OL* 13:163-186.
- 1977 The internal relationships of Rukai. *BIHP* 48/1:1-90.
- 1978 A comparative vocabulary of Saisiyat dialects. *BIHP* 49/2:133-199.
- NOTHOFER, Bernd
- 1975 *The reconstruction of Proto-Malayo-Javanic*. *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde* 73. The Hague: Martinus Nijhoff.
- REID, Lawrence A.
- 1971 *Philippine minor languages; word lists and phonologies*. *OL Special Publication* 8. Honolulu: University of Hawaii Press.
- 1979 Evidence for Proto-Philippine nominative marking. *Philippine Journal of Linguistics* 10:1-20.

SCOTT, N.C.

- 1956 *A dictionary of Sea Dayak*. London: School of Oriental and African Studies, University of London.

SNEDDON, James N.

- 1978 *Proto-Minahasan: phonology, morphology and wordlist*. PL, B-54.

TING, Pang-hsin

- 1978 Reconstruction of Proto-Puyuma phonology. *BIHP* 49/3:321-392.
[English abstract 391-392.]

TSUCHIDA, Shigeru

- 1967 Itbayaten reflexes of Proto-Austronesian phonemes. Yale. Typescript.
1969 Word list of three Bunun dialects. Yale. Typescript.
1976 *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No.5. Tokyo.

WILKINSON, R.J.

- 1959 *A Malay-English dictionary*. London: Macmillan.

ZORC, R. David

- 1971 Proto-Philippine finder list. Cornell University. Typescript.
1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. PL, C-44.
1978 Proto-Philippine word accent: innovation or Proto-Hesperonesian retention? In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 67-119. PL, C-61.
1979 Core etymological dictionary of Filipino. Fascicle 1. Manila: De La Salle University.

THE EVOLUTION OF FOCUS IN AUSTRONESIAN

Stanley Starosta, Andrew K. Pawley and Lawrence A. Reid

1.0. INTRODUCTION

1.1. The problem

In this paper,¹ we will attempt to reconstruct the features of Proto-Austronesian morphology and syntax which gave rise to the focus systems exhibited by modern Philippine languages. In order to approach this problem, it will be necessary to consider the following questions:

1) What is the grammatical structure of sentences showing 'verbal focus' in Philippine languages? And in particular, what is their synchronic and diachronic relation to nominalisations which show affixes cognate with the verbal focus affixes? We need to have a reasonably clear idea of the endpoint of an evolutionary sequence before we can reconstruct the stages that led up to it.

2) Do the focus systems of Philippine languages represent a retention from Proto-Austronesian or an innovation? What kind of case marking system can we reconstruct for the proto-language which will allow us to provide plausible accounts of how a single original system could evolve into the Oceanic object focus system in one area and the Philippine subject-focus system in another?

An attempt to answer 2) will require consideration of such specific questions as:

3) What are the higher-order subgroups within Austronesian? The position we take on this question of course will determine which combinations of languages will count as adequate witnesses for reconstructing a morphological or syntactic feature all the way back to the proto-language.

4) What is the current distribution of Philippine-style focus systems by geographic regions and within subgroups of Austronesian languages? This will determine how far back we can reconstruct this syntactic property.

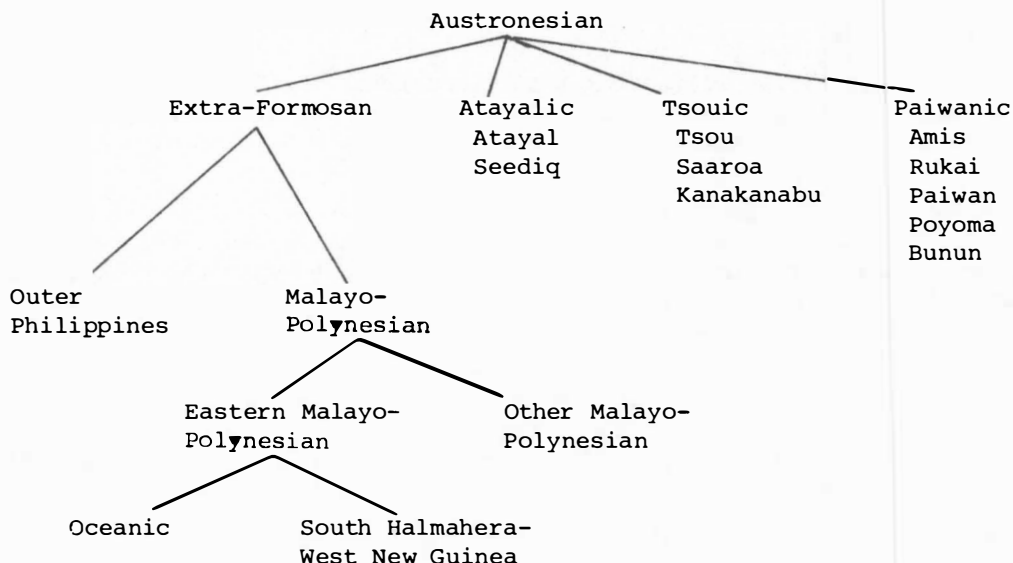
5) How likely is it for two languages to have developed a Philippine-style focus system independently? To answer this question, we have to make assumptions about what kinds of syntactic changes are possible and likely. By rejecting excessively abstract syntactic representations and arbitrary analyses and formulating our solution within the narrow constraints of lexicase (Starosta 1979), we eliminate a large class of conceivable but ad hoc and unmotivated analyses, and come up with an account of the evolution of focus which requires no hypothetical stages having properties which cannot be directly observed in the 'surface structures' of modern human languages.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 145-170. *Pacific Linguistics*, C-75, 1982.

© Stanley Starosta, Andrew K. Pawley and Lawrence A. Reid

1.2. Subgrouping assumptions

In this paper, we will assume the correctness of Dahl's (1973) and Blust's (1977:2) recent hypotheses about the first-order subgroups of Austronesian. Reid argues further in his paper for this conference (see pp.201-216 in this volume) that at least the Northern Philippine languages constitute a primary subgroup, called Outer Philippines, of these extra-Formosan languages, but the correctness of this claim does not affect the validity of our arguments in this paper. The subgrouping assumptions within which we are working are illustrated in the following tree diagram.



2.0. PROTO-AUSTRONESIAN: OBJECT FOCUS, SUBJECT FOCUS, OR NONE OF THE ABOVE?

In order to say anything sensible about where 'focus' came from, we have to know 1) what focus is, and 2) whether words marked by 'focus' affixes in Philippine languages are nouns or verbs. In this paper, we will use the term focus to refer to a system of verbal affixes used to indicate the case relation of the subject of a sentence.

Most modern linguists working on Philippine languages, from Bloomfield and Blake on up to recent studies by the Summer Institute of Linguistics people and lexicase grammarians such as Harmon and De Guzman, have assumed almost without question that 'focussed' words are verbs. The correctness of this conclusion is however not immediately obvious. Cecilio Lopez (1941) and A. Capell (1964) both consider all Philippine 'passive' verbs to be verbal nouns. Capell based his conclusion essentially on the fact that agents in these constructions appear in the Genitive case form.

Similar conclusions have been drawn for analogous reasons about passive verbs in Atayal (Egerod 1966:346) and Toba Batak (Van der Tuuk 1971), and about one of two types of 'passive' construction in Rukai (Li 1973:202-211). Ferrell (1974:5-8) raises this possibility for Paiwan, but rejects it for semantic and pedagogical reasons, although he concedes that his decision is based on a

'lingua-centric view'. McKaughan (1962:49, note 8) also rejects a nominal analysis because nouns should not be marked for tense, aspect, and voice. Similarly, Schachter and Otnes say that all basic Tagalog sentences are essentially equational in nature (1972:62; cf. p.117; cf. also Dahl 1973:117-118). However, they treat basic sentences as verbal because they find a verbal treatment to be more 'convenient'.

We don't find the arguments in the preceding paragraph very persuasive. 'Convenience', pedagogical or otherwise, has no status as a scientific criterion, and the use of the presence of 'tense, aspect, or voice' to exclude a nominal interpretation is circular, since that is what we are trying to decide in the first place. As for aspect, Pawley and Reid (1979:109) note that focussed and aspect-marked words are frequently used as common nouns, and that some focussed forms can *only* occur as nouns.

We will take the position here that, while many clauses in languages such as Tagalog, Amis, or Ilokano can be given neat and satisfying analyses as binary NP - NP cleft sentence structures, some can't, due to the fact that a full NP subject occasionally intervenes between the lexical head of the predicate and the other actants of the sentence.

There are two prime candidates for the reconstruction of the Proto-Austronesian case-marking system:

1) the Proto-Oceanic system, in which the verb carries a suffix (*-i or *-akin) to indicate the case relation of the direct object: (*-i for Locus, *-akin for Instrument or Referent), and

2) a Philippine-style focus mechanism utilising the verbal affixes *mu-/-um-, *ni-/-in-, *-en, *-ana, and *iSi- (not cognate with the Oceanic suffixes *-i and *-akin) to indicate the case relation of the grammatical subject rather than the object, with the affixes *mu-/-um- marking verbs with Agent subjects, *ni-/-in-, *-en, *-ana, and *iSi- marking Patients, *-ana marking Locus, and *iSi- marking Instrument or Referent.

Each of these candidates has had its supporters. William Foley (1976) has claimed that Proto-Austronesian case marking must have been similar to that of 'classical' Oceanic languages such as Fijian. Dahl (1973) and Wolff (1973) however, both concluded that PAN should be reconstructed with at least the four morphological focus or voice contrasts marked by reflexes of *mu-/-um-, *ni-/-in-, *-en, *-ana, and *iSi- that are generally present in modern Philippine languages. Similarly, Pawley and Reid (1979) argue that Philippine-style focus systems are a retention from PAN, in their essentials, and that the Proto-Austronesian focus system has decayed, to a lesser or greater extent, in languages outside of a region comprising the Philippines and certain contiguous regions of Indonesia and Formosa. Thus, the common possession of a focus system should not count as evidence for treating Philippine languages as a subgroup.

Pawley and Reid (1979:111) also noted that "... the use of verb stems plus non-Actor focus affixes as nouns is clearly PAN. The nominal uses are found throughout Philippine type subgroups as well as in Oceanic and Toba Batak of Sumatra, and their PAN status can hardly be questioned". However, they conclude that it is probably necessary to reconstruct PAN verbal passive constructions involving the same set of affixes, and that the verbal usage preceded the use of the affixes as nominalisers.

They derive the Oceanic case-marking type from an intermediate stage of development similar to that persisting in Toba Batak. The Batak system combines

features of both the Philippine and Oceanic systems of case-marking and focus, e.g. showing both subject-focus affixes on the verb in passive sentences (cognate with those of Philippine languages) and object-focus suffixes on the verb in active sentences (cognate with those found in Oceanic languages). Pawley and Reid tentatively suggest that PAN may have been like Toba Batak in these respects.

In the present paper, we will argue that *-en, *ni-/-in-, *-ana, *iSi-, and possibly *mu-/-um- were all noun-deriving affixes in PAN, as they still are to a large extent in the modern languages outside the Philippine area, and that they have in fact retained this function to a previously unrecognised extent even within the Philippine language group. We argue further that Austronesian nominalisations in *-en, *ni-/-in-, *-ana, *iSi-, and possibly *mu-/-um- did not develop from original passive constructions, as concluded by Dahl (1973), Wolff (1979), and Pawley and Reid (1979), but rather that the nominalising function was the original one, and that the passive and verbal focus uses of these affixes in Philippine languages are a secondary development. That is, verbal focus in Proto-Austronesian was at most an incipient mechanism that was later elaborated and developed by the languages of the Philippines and some languages of Borneo and the Celebes.

If this argument is correct, then the possession of a well-developed verbal focus system becomes potential evidence for subgrouping, depending on how likely it would be for focus to come into existence independently in separate subgroups, and on how likely a focus system could spread as an areal feature among separate subgroups.

Our arguments for this hypothesis include the following:

- 1) Throughout the Austronesian family, but especially in those languages which show verbal focus, the person marker forms for the agents of passive verbs are the same as the genitive pronouns marking the possessors of underived nouns, and contrast with the other sets of person markers.
- 2) The reflexes of the 'focus affixes' mentioned above outside the Philippines are very largely nominal derivational affixes, and even in languages such as Malagasy and Toba Batak, it now appears as if many constructions previously analysed as verbal may turn out to be amenable to a nominal construal, just as their counterparts in Philippine languages have turned out to be.
- 3) The odd patterns of focus affixation in verbs, with some case inflections being suffixed (*-en, *-ana), some prefixed (*iSi-, *mu-, *ni-), and some infixes (*-um-, *-in-), suggests that focus paradigms are the result of the welding together of originally disparate elements, the originals in most cases being most plausibly derived from nominalising morphemes.
- 4) While deriving the nominal forms from passive constructions can only be done with ad hoc and unmotivated transformational rules, we have found a plausible way to derive verbal focus constructions from nominal ones which involves only a simple reinterpretation of isomorphic clauses and relabelling of several crucial nodes.

3.0. PROTO-AUSTRONESIAN SYNTAX

At this point, it is convenient to give a brief sketch of PAN sentence structure as we reconstruct it.

Proto-Austronesian was probably a verb-initial split ergative language

like Amis or Palauan, with ergative Agents and possessors both marked by the same Genitive case form, a common feature of ergative syntax. Tense, aspect, negation, and various adverbial notions such as manner were carried by a small class of verbs which, like 'auxiliary' verbs generally, were the grammatical main verbs, the lexical heads of their sentences, with other verbs occurring in sentences under the 'auxiliaries'. Nominative and Genitive clitic pronouns were 'attracted' to the syntactic heads of the main sentence.

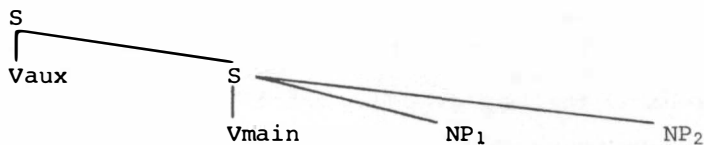
The normal position for the Genitive Agent of an ergative clause was immediately following the head verb of its clause (possibly with one or more intervening clitic pronouns or adverbs), since otherwise it could be interpreted as a Genitive attribute of the noun preceding it.

Grammatical subjects were obligatorily definite, that is, assumed by the speaker to be identifiable by the hearer from the linguistic or extralinguistic context. All of these properties can be observed in modern languages such as Tagalog (McFarland 1978), Amis, and Tsou, and so can be reconstructed for PAN.

PAN was a strongly noun-oriented language, with a high percentage of nominalisation strategies. The affixes *-en, *ni-/-in-, *-ana, *iSi-, and possibly *mu-/-um- functioned to derive nouns from verbs, with only *-en possibly having begun to function to derive verbs as well as nouns.

4.0. AUXILIARIES AS MAIN VERBS IN PAN

PAN must have had an extensive set of auxiliary verbs, a set which almost certainly included not only words marking tense or aspect, but also logical and existential negators and certain kinds of 'adverbs' denoting manner and instrumentality (cf. Starosta 1974:300-301, 315, 319, 333-334, 347-349 and Chen 1982). On the basis of evidence from languages throughout the Austronesian family, we can conclude that these elements were in fact grammatically verbs, and that in spite of the implications of the term 'auxiliary', they were syntactically the grammatical heads of their constructions, with the so-called 'main verbs' being syntactically embedded under the 'auxiliaries' as sentential complements. That is, the appropriate analysis for auxiliary verbs in Austronesian languages is something like:



The generalisations that can be captured by this analysis include the following:

1) Word order: Instead of saying that the initial element in the sentence (assuming no topic is present) is a predicate nominative or a V unless an Aux is present, we simply say, unless a topic is present, the initial element in every clause is the head of the predication: NP, PP, or V, period.

2) Clitic placement: Instead of stating that clitic pronouns and clitic adverbs are attracted to the NP predicate or main verb unless one of a set of preverbal elements is present, in which case the clitic for some unknown reason precedes instead of following the 'main' verb (see, for example,

Schachter and Otnes' discussion (1972) of the various classes of elements that obligatorily or optionally precede clitics), we can state simply that clitics are attracted to the lexical heads of their constructions, whether NP, PP, or S.

3) Dependent verb inflections: Certain languages such as Kagayanen (Harmon 1977:100ff), Seediq (Asai 1953:28), Manobo (Elkins 1971, Morey 1964, DuBois 1976), Maranao (McKaughan 1958), Samareño (Wolff 1973:82, 86), and Atayal (Egerod 1966) have a set of verbal inflections that occur only in imperatives or when the verb is either embedded under another verb or follows certain elements of a set of auxiliary words marking aspect, negation, etc. Under the analysis we propose here, we need only state that verbs must appear in dependent inflected forms either in imperatives or when they are dependent, that is, when they are embedded under higher verbs. This aspect of our analysis becomes very important in accounting for the change from PAN to languages of the Oceanic type.

5.0. PAN NON-VERBAL CLAUSES

Proto-Austronesian non-verbal clauses were composed of an initial predicate noun phrase or prepositional phrase followed by a grammatical subject and optional outer circumstantial actants such as Time and Place. There was no copula in such sentences. Non-verbal clauses, like verbal ones, could be embedded under auxiliary verbs marking tense, aspect, and negation.

Predicate nominative sentences were either descriptive, with indefinite predicates, or identificational, with definite predicates.

Descriptive predicate nominatives did not have their own referents. Rather, they added information about the subject of the clause. Except for having the basic internal structure of a Noun Phrase, they were essentially identical in their syntactic properties to stative verbs, even to the point of allowing nominative clitic pronouns to attach to the head predicate noun. PP predicates also supported clitics.

As in the case of verbal clauses, the nominative clitic was obligatory in main clauses when the implied subject was first or second person. There was no overt third person singular nominative clitic, and probably no overt third person plural either.

The second type of predicate nominative sentence, or identificational, took two definite NP's with independently registered referents and identified them with each other. This type, too, is widely attested in Philippine and Formosan languages, although it is probably far less frequent than the descriptive type.

The equational sentence type was almost certainly very frequent, as it continues to be in Paiwanic and Philippine languages, and as will be shown below, it had a crucial role to play in the evolution of verbal focus inflections from nominalising derivational affixes.

6.0. PAN AS A VERB-INITIAL SPLIT ERGATIVE LANGUAGE

We assume that Proto-Austronesian was verb initial because this is the usual word order in Philippine and Formosan languages as well as in such languages as Toba Batak and Merina (cf. also Wolff 1979:164). Emphatic, contrastive, or presupposed NP's or adverbials could appear as preverbal topics, immediately

followed by an intonation break.

The claim that PAN was a split ergative language is based on the following considerations:

1) Within the lexicase framework, an ergative language is defined as one in which the grammatical subject is always in the Patient case relation. A split ergative language is one in which the unmarked subject choice is Patient, but which has one or more classes of derived verbs which choose their grammatical subjects according to Fillmore's (accusative) Subject Choice Hierarchy: Agent first, else Instrument or Correspondent, else Patient (using lexicase labels for the case relations).

2) A number of languages from different primary Austronesian subgroups, including Tongan, Samoan, Ilokano, Palauan, Chamorro, Toba Batak, Paiwan, Amis, and Tagalog (cf. De Guzman 1978:199) are ergative or split ergative in the sense of 1) above.

3) In the split ergative languages, the ergative verb stems are often less marked than the accusative ones, and the completely unmarked 'root-stems' (De Guzman 1978:199) are always ergative in languages such as Tagalog, Kagayanen (Harmon 1977:111, Table 6), and Toba Batak (Van der Tuuk 1971:85, 98) where 'simple passives' consist of a bare stem, while 'active' transitive verbs are derived (cf. Mulder and Schwarz 1981:250 on Bilaan). That is, Toba Batak 'simple passives' are grammatically ergative, since the unmarked subject is the Patient rather than the Agent.

Even in languages which have drifted off in an accusative direction, non-subject Agents tend to be marked by the same case form as possessors, a typically ergative characteristic, and derived but otherwise unmarked *pa- causative stems tend to retain their original ergative properties. Thus *pa- causatives in Kapampangan (Mirikitani 1972:79), Kagayanen (Harmon 1977:111), Tsou (Tung 1964:225), Tagalog (De Guzman 1978:339), Seediq, and to some extent in Atayal (Egerod 1965:267) and Bunun have Agents in their case frames but allow only Patients as grammatical subjects unless further derived.

4) Linguists such as Ceña (1977) and De Guzman (1979) have pointed out Tagalog's tendency to 'Patient Primacy', the typically ergative inclination to give preference to Patients in subject choice, morphological marking, etc. This tendency is reflected for example in the fact that if a Tagalog sentence refers to a Patient and an Agent which are both definite, only the Patient can be chosen as the grammatical subject.

5) Finally, note that the Agents of imperatives in Austronesian languages are typically non-subjects. This is the case for example in languages such as Maori (Clark 1973:577), Hawaiian, Betawi (Ikranagara 1975:124), and Formosan languages such as Tsou (Tung 1964:84), Bunun, and Amis. The fact that imperatives in languages such as Seediq (Asai 1953:56) preserve reflexes of the original derivational suffixes *-i or *-a even when, as is the case in for example Amis, Rukai, Saisiyat, and Bunun, these have been lost elsewhere in the language, and that archaic forms of the verb root can occur in imperatives (e.g. Bunun *koni 'eat'*, as compared with the regular form *ma'un*), provide additional support for our contention that Patient-subject imperatives were a feature of the ergative proto-language.

7.0. THE STRUCTURE OF NOUN PHRASES

7.1. Heads and attributes

Proto-Austronesian Noun Phrases were composed of a head noun optionally followed by one or more NP attributes, or possibly by a verbal relative clause. NP attributes following noun heads were either Locative (as in English 'the woman in the pool'), Genitive (as in 'the name of the game'), or appositional (as in 'my son, the hunter').

7.2. Adjectives and demonstratives as nouns

The \bar{X} convention as interpreted within the lexicase framework (Starosta 1979:60) requires that the lexical head of a Noun Phrase be a noun. However, it should be noted that the lexical items that must be classified as nouns according to syntactic criteria in Proto-Austronesian and in many of the descendants often correspond to adjectives or demonstrative determiners in English translations, and this correspondence has unfortunately influenced the synchronic analyses of many Austronesian languages, where it has been assumed without question or justification that a determiner or adjective in the English translation is necessary and sufficient grounds for postulating a determiner or adjective in the language being analysed. For PAN, the only determiners we presently reconstruct are the Genitive $*i/*ni$ (cf. Reid 1981) and a personal Nominative article $*si$.

8.0. VERBAL DERIVATION WITH $*-i$ AND $*-aken$

In addition to an inventory of unmarked and $*pa-$ causative ergative verb stems, Proto-Austronesian also had derived verb stems suffixed by $*-i$ and $*-a$, and perhaps other elements such as $*-aken$ or $*-neni$. These suffixes were homophonous with synchronically co-existing prepositions $*i$, $*a$, $*aken$, and possibly others, and were diachronically derived by a process of preposition capture of the sort that operates in German (ausreissen 'tear out' vs reissen 'tear'), Latin (extrahō 'draw out, extract' vs trahō 'draw, drag'), or Mandarin Chinese (jìgěi 'send to' vs jì 'send'; Hou 1979:79). $*i$ and $*aken$ had two functions: recentralisation and definite marking.

In a lexicase grammar, the Patient case is the fundamental case relation. Every verb, with the occasional exception of ambient or meteorological verbs such as (in some languages) '(It) is-hot' or '(It) is-snowing' has at least a Patient in its case frame, and this Patient is viewed as the central element in the action or situation designated by the verb. Many languages, however, have a mechanism for varying the 'perspective' (Fillmore 1977:72-79) of a given verb stem, and in lexicase, this means treating some other actant associated with the verb root as the Patient, and either reinterpreting the original Patient as some other case relation or excluding it altogether from the case frame. To take several examples from English:

John climbed over the mountain.	
Nom	Lcv
PAT	LOC

John climbed the mountain.	
Nom	Acc
AGT	LOC

Joe Bloggs fought with the champion.

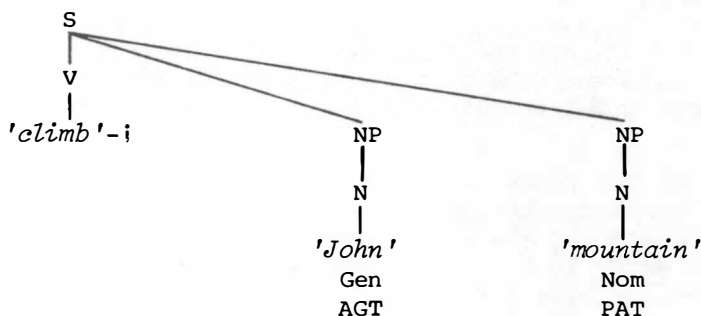
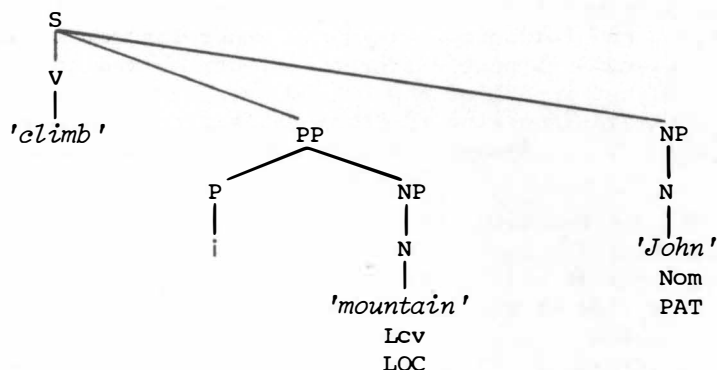
Nom	Lcv
PAT	LOC

Joe Bloggs fought the champion.

Nom	Acc
AGT	PAT

In these examples, actants marked by an oblique case-form preposition are reinterpreted as direct objects (Accusative Patients), and in many languages, this derivation process also involves the retention of the original preposition as an independent adverb or as a fused affix of the verb itself.

The derivational process which reinterprets a different case relation as Patient can be referred to as 'recentralisation', since in effect it places a new situational role in the perceptual centre of the stage. In PAN, this process was quite productive, and exactly as in German, the derived verb stems were marked by affixes derived from the prepositions originally captured from the oblique actants that were 'centralised'. One difference between German and Proto-Austronesian, of course, was that PAN was a verb-initial language, so that the P's followed the verbs and were suffixed, instead of being prefixed as in German. The other relevant difference is that PAN was ergative, and an ergative language is one in which the Patient is always the grammatical subject. This means that when a Locus actant, say, was reinterpreted as Patient and lost its *i preposition to the verb, it became the grammatical subject of the new verb, and the new *-i suffix on the verb became a marker indicating that the subject of the sentence was situationally locational. This is depicted in the following schematic example:



Note that the lexicase approach to this phenomenon involves a fundamental change in case relation, thereby providing an explanation for the difference in semantic interpretation. On the other hand, a relational grammar account of these data, for example, would involve only a difference in 'grammatical relation', a category whose semantic implications are unclear.

As mentioned at the beginning of this section, recentralisation was only one of the functions of the **-i/*-aken* verbal derivation process. The other was definitivisation. As mentioned previously, PAN grammatical subjects were obligatorily definite. Thus a simple way to mark an actant as definite was to recentralise it, that is, to reinterpret it as a Patient, thereby making it the grammatical subject and thus grammatically definite.

Clearly, what we have reconstructed here is something very much like what is called 'focus' in Philippine linguistics: a system of verbal affixation which allows different actants to be placed in subject position, thereby marking them as definite, and which signals the presence of a particular situational role associated with the subject. That is, we are very close to the position taken earlier by Wolff, Dahl, and Pawley and Reid. The difference, of course, is that we don't think the 'focus' system of PAN was marked by the usual Philippine-style **-en*, **i-*, **-an*, or **-in-/ni-* affixes. Rather, at the beginning at least, it was implemented by the elements ancestral to the Oceanic transitive markers, a view which in this respect at least is closer to Foley's position (Foley 1976: 214ff).

The Proto-Austronesian **i/*aken* verbal derivation system has its most striking reflection in Oceanic, but it is by no means limited to this subgroup, and the actual affixes we reconstruct are reflected not only in Oceanic, but in fact in Chamorro (Topping 1973), Toba Batak, Bahasa Indonesia, Bisayan languages, Inibaloi, Marinduque Tagalog, and in all three Formosan subgroups, Atayalic (Atayal, Seediq), Tsouic (Tsou), and Paiwanic (Amis). In other languages such as Tsou (Tung 1964:224-225) the suffixes have different effects depending on the stem to which they are attached, but in all of these languages, the function of the **i* and **aken* reflexes are similar: marking the Patient as a derived one associated with some other non-Patient grammatical role implied by the source verb stem.

At this stage of our work, one question remains open: the relation between **a* and **aken*. There is an asymmetry in our reconstruction, because the suffixes involved in the recentralising derivational process just discussed, especially as reflected in the languages outside of the Philippines and Formosa, are **i* and **aken*, whereas the affixes reflected in the dependent verb forms in Formosa and the Philippines are **i* and **a*. The **i* in these languages behaves quite regularly, but while the dependent verb suffix *-a* in Atayal corresponds grammatically to **aken*, marking the centralisation of peripheral 'accessory' case roles (Egerod 1966:353) the *-a* in Tsou and in dependent and imperative verbs in Philippine languages corresponds to the OF **-en*, not the Referential **iSi-* as it should if it corresponded grammatically to **aken*. (Tsou does have a suffix *-(n)eni* which corresponds in function to **aken*, but there seems to be no way to link these two forms historically.)

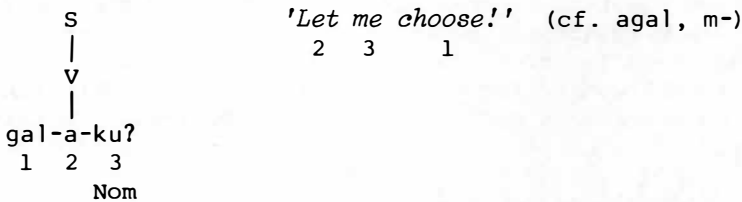
Thus the *-a* in Formosan and Philippine languages usually marks 'Object Focus' rather than 'Referential' or 'Accessory' focus. It seems that both suffixes indicated transitivisation, but that the **-a* functioned to derive transitive verbs from intransitives by adding an Agent to the case frame, whereas the **-i* indicated that a transitive verb had been derived by 'centralising' the original Locus (reinterpreting it as Patient), thus requiring the original Patient to assume the Agent role.

*i

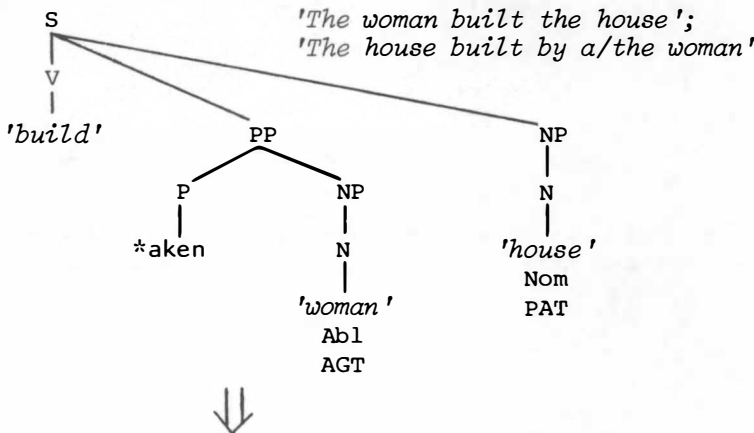
The reconstruction of *i is amply justified by its widespread reflexes throughout Austronesian (cf. Dahl 1973:119). This preposition, which was the source of the *-i suffix, was a general non-terminus Locative preposition which marked Locus and Correspondent. In Philippine languages such as Tagalog, and in the Formosan language Amis, it is possible to have situational objects appear as non-subjects in certain constructions, but the case form in which they appear depends on the class of noun: common nouns are Genitive in Tagalog or Accusative in Amis, but personal nouns are marked as Locative, which in Amis involves the preposition i. If this feature is reconstructible for PAN (which seems rather doubtful at the moment in the absence of evidence from the other Formosan languages), this Locative *i could conceivably be the source of the -i which marks transitive verbs in general in Oceanic.

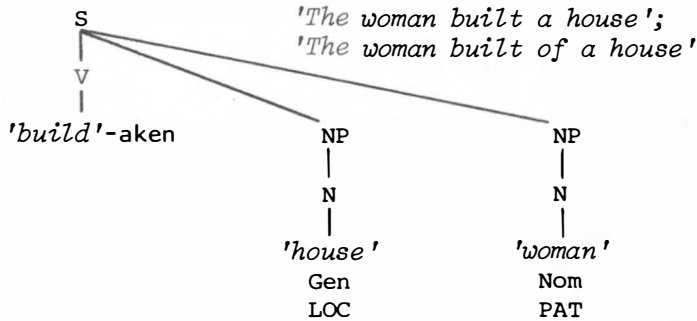
*aken

The suffix and preposition *akin is reconstructible for Proto-Oceanic (Pawley and Reid 1979) and it has cognates for example in Wolio (Anceaux 1952) and Bahasa Indonesia. This element marked a general terminus Locative case form, and when captured in a recentralising derivation, it added a terminus component of meaning to the derived verb. As a preposition, *aken probably marked Agent/Instrument as well as (comitative) Locus case relations. Thus we find -a as a marker of 'Agent Focus' in subordinate clauses in Atayal (Egerod 1966:353):



and -kan as a causative affix in Indonesian (MacDonald and Dardjowidjojo 1967:90). In both cases, the suffix represents an oblique preposition captured from a non-subject Agent actant in an ergative clause in the process of recentralisation, as in the following schematic diagram:

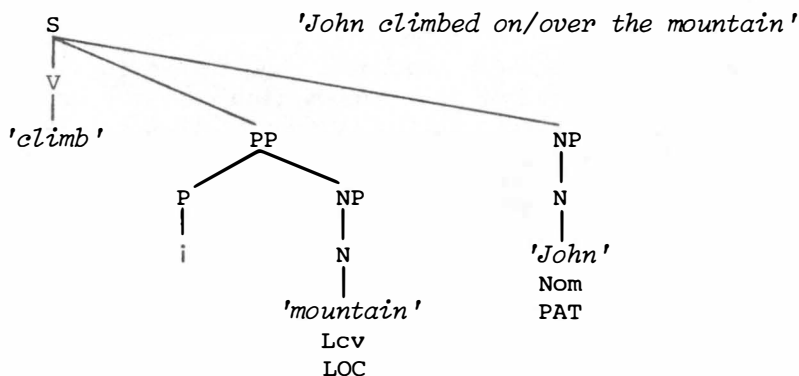




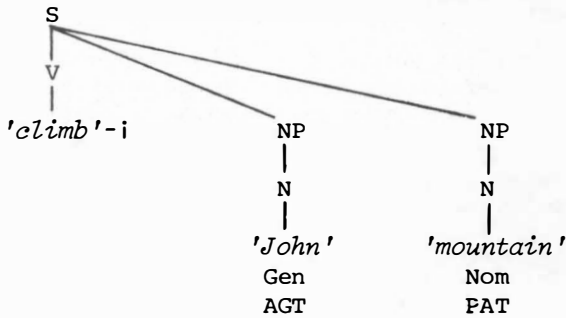
As mentioned above, however, it is not yet clear whether this form can be reconstructed in its verb-deriving function all the way back to PAN. The verb-deriving process itself is certainly reconstructible, however, but the most common exponent of it in Philippine and Formosan languages is a reflex of *a rather than of *aken, as in the Bunun imperative qanup-a 'Hunt!'.

9.0. THE ORIGIN OF PHILIPPINE VERBAL FOCUS

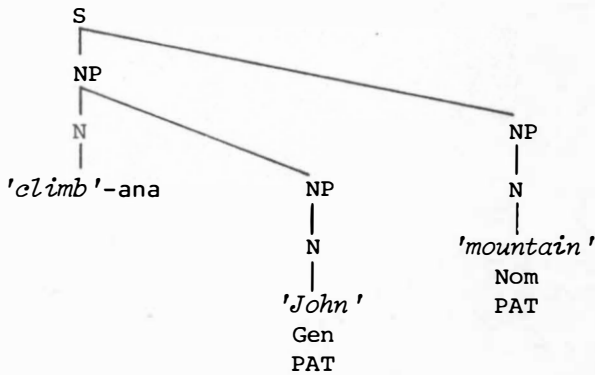
The cleft-sentence constructions are interesting for our purposes because they provided an alternative strategy for 'recentralisation', that is, of recasting some actant in a non-subject case relation as the Patient, the perceptual centre of the action or situation and the presupposed element in the predication. This is accomplished by taking one non-subject non-Patient actant and making it the subject of a descriptive equational predication. Since it is the subject, it must be definite, and since equational predicators are one-place predicators and thus have only one case-relation slot available, that slot must be filled by a Patient, since Patient is obligatory for every (finite) clause. Thus the subject of an equational sentence is a definite nominative Patient. To cite the example given earlier:



One way to focus on 'mountain' in this structure, of course, is the method discussed in the preceding section: make it the subject of an *-i-suffixed ergative verb:

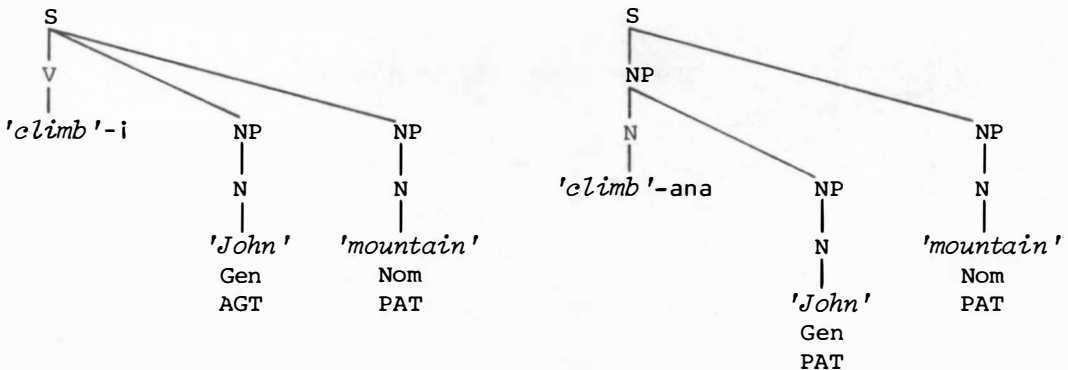


Instead of this, however, we could make it the subject of a nominalised verb, using the deverbal nominalising suffix **-ana* 'place where', e.g.



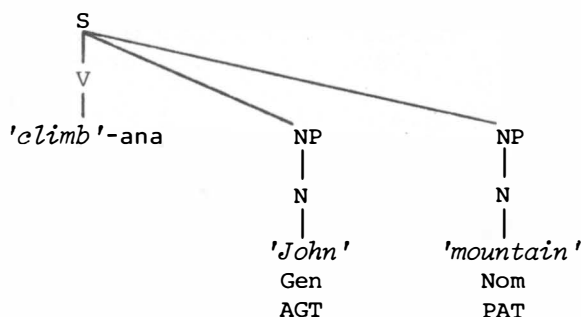
The end effect of the **-ana* nominalisation and the **-i* verb derivation are then in effect the same: the Locus actant 'mountain' is converted to a Patient and made the definite subject of the clause; that is, to use Philippinist terminology, it is 'focussed', with the suffix *-i* on the verb and the suffix *-ana* on the nominalised noun both serving to mark the Patient subject as associated with a situational location.

What we have in PAN, then, is two alternative and competing strategies for focussing non-subject actants:

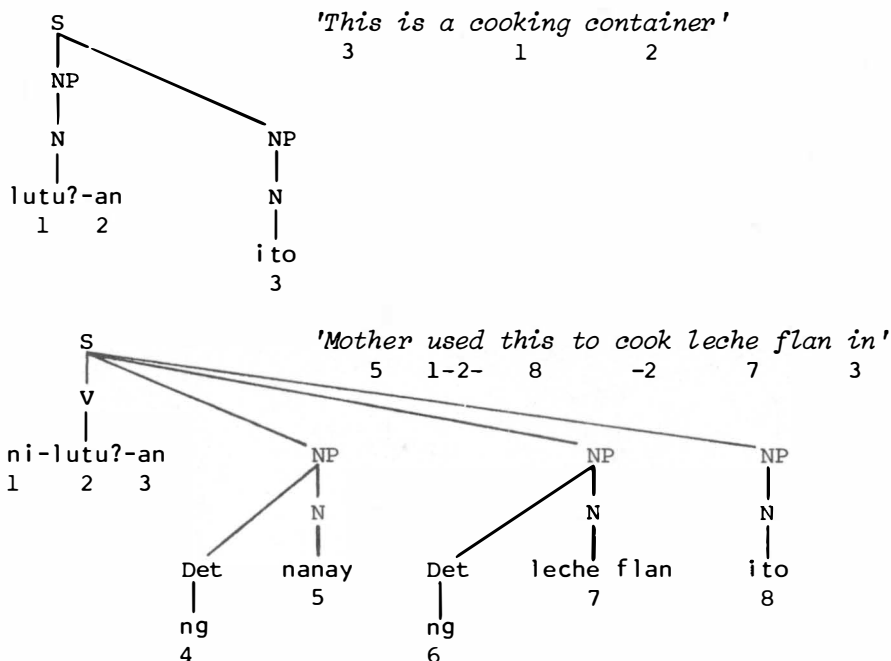


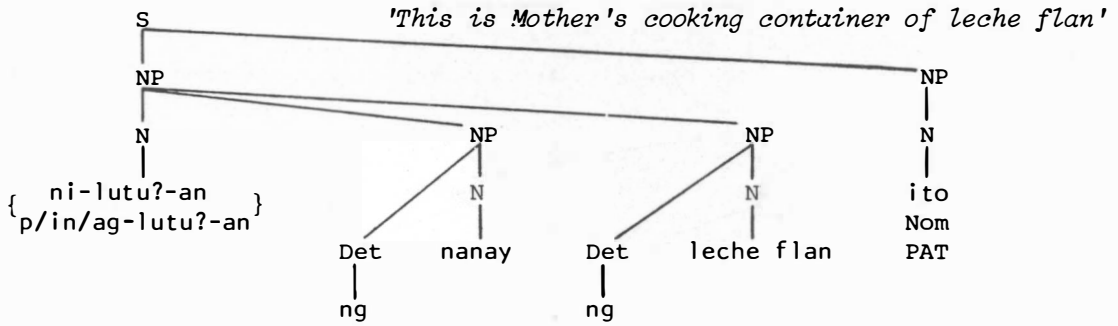
9.1. The reinterpretation of PAN cleft sentences as verbal

When we notice that these two structures match up word by word and case form by case form, it is easy to see how the next stage of the development of Philippine-style verbal focus came about: some (though certainly not all) of the nominal structures were reinterpreted as verbal ones. The simplest way to visualise this is to say that the sequence 'climb'-i in the verbal structure was replaced by the phonological sequence 'climb'-ana without changing the lexical matrix in any other way. The result was a derived ergative structure which is superficially identical to a Philippine-style focus, and differs only in that the subject is still Patient:



Note that this derivation-by-reinterpretation did not alter the source noun entry 'climb'-ana in any way, so that both 'climb'-ana entries coexisted in the lexicon. This situation continues in Tagalog, for example, where a PAN-style deverbal local noun coexists with a homophonous LF verb, and sometimes with secondary deverbal nouns as well (De Guzman, personal communication):





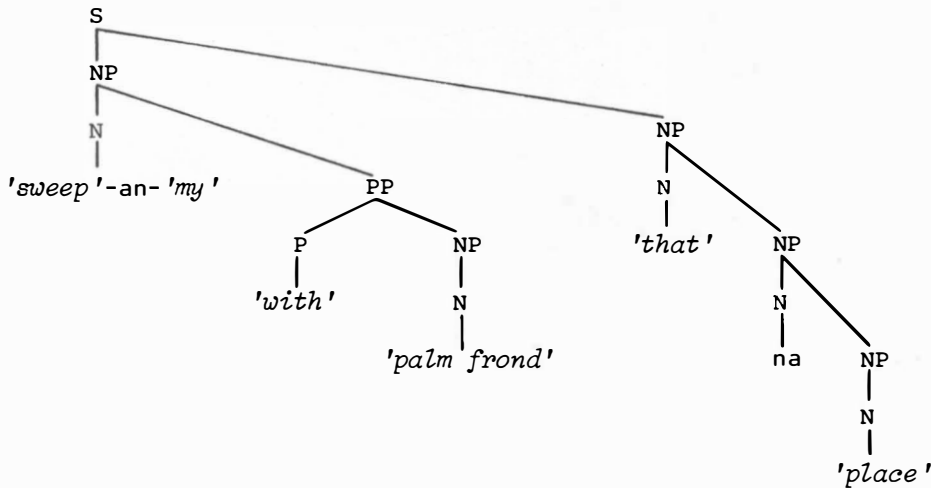
9.2. Cleft sentences with *mu-/um-, *-en, *ana, *iSi and *ni-/in-nominalised predicates

In order for the derivation of verbal focus to have worked as we hypothesise, there must have been a range of nominalisation affixes matching the case roles to be focussed. The ones primarily involved were:

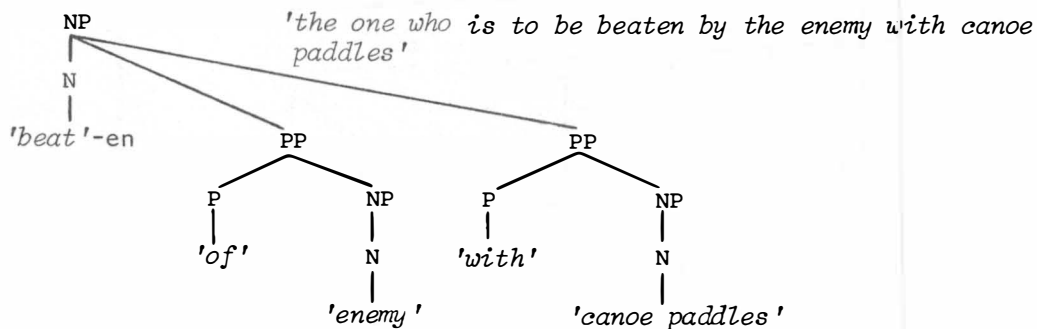
- *mu-/um- 'Actor focus'
- *-en 'Goal focus'
- *-ana 'Referential focus' (Dahl 1973 reconstructs *-an)
- *iSi- 'Instrumental focus' (Dahl 1973 reconstructs *Si-)
- *ni-/in- 'Perfective'

Deverbal nouns derived from verbs using these affixes occurred in descriptive equational predicates of the sort illustrated by the following schematic examples:

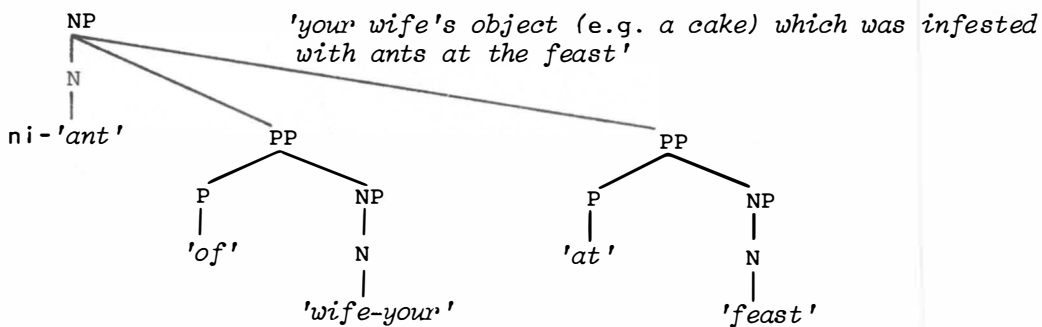
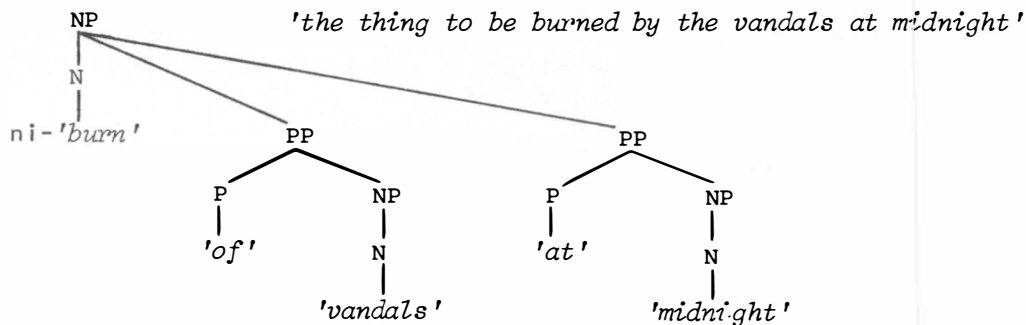
*-ana 'place of V-ing'



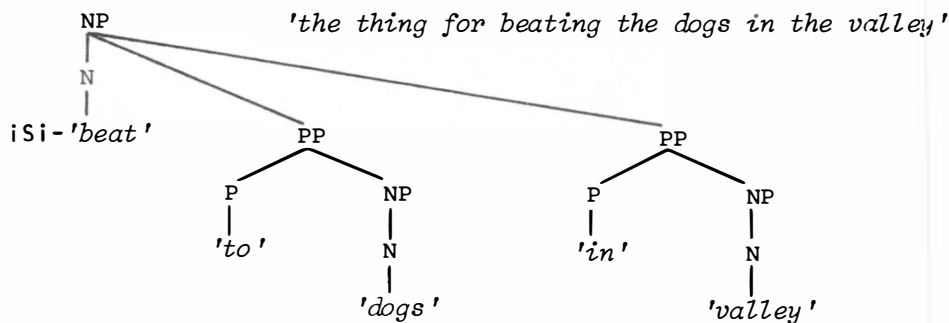
*-en 'the N to be V-ed'



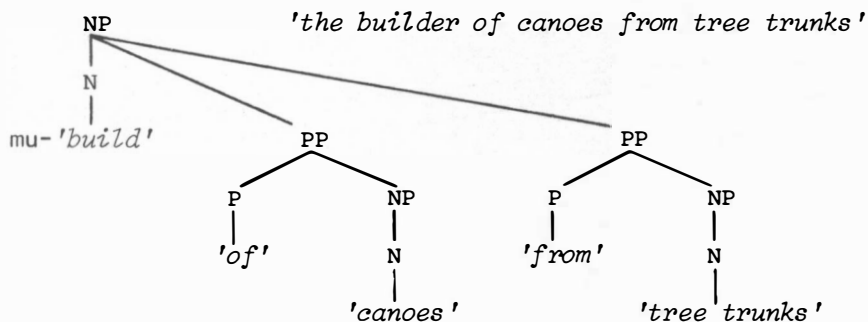
*ni-/in- 'the N affected by V-ing'



*iSi- 'thing for V-ing or for N'

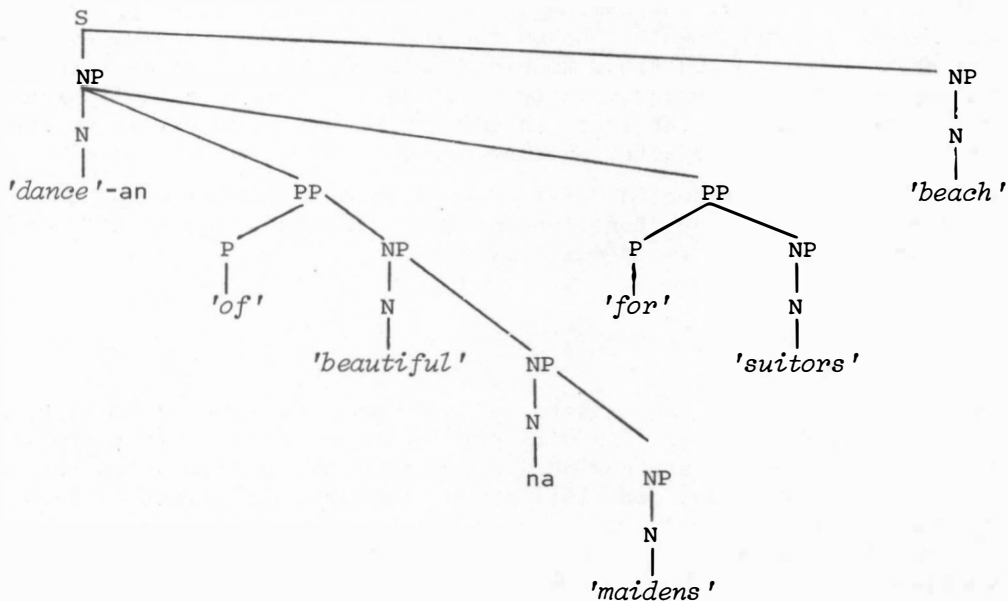


*mu-/-um- 'one who V's'



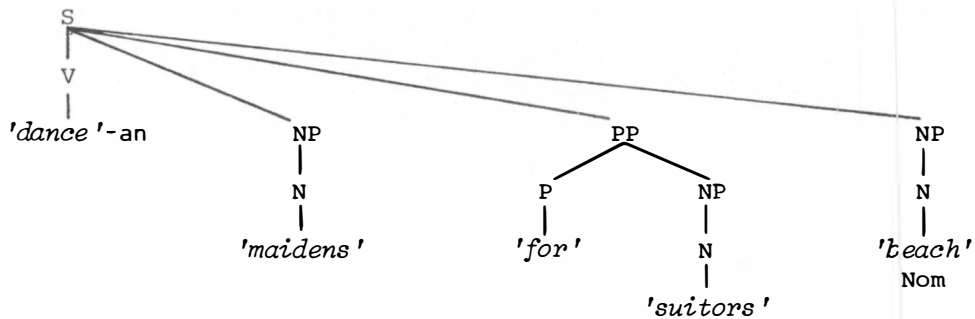
During the transition period, isomorphous structures were internally represented by some speakers as nominals and by other speakers as verbal. The nominal constructions of course were always subject final (allowing for final outer Time and Place actants), since the grammatical subject of an equational sentence is one of two immediate constituents in the sentence, and so cannot be in the middle of the other immediate constituent:

Nominal



For those speakers with corresponding isomorphous verbal constructions, though, this constraint on constituent order would not have to be absolute, since the verbal structures were not limited to binary branching constructions:

Verbal



This means that as soon as the verbal speakers shifted the subject into a non-final position, the nominalisation speakers were placed on notice that something was different, and were given the crucial clue they needed to reinterpret at least some of their cleft constructions as verbal.

This would help to explain why it is that in Philippine and Formosan languages, and in many Indonesian languages as well, relative clauses are exclusively nominal constructions: since the grammatical subject of the relative clause was co-referential with the head N of the NP and thus omitted ('deleted') for both verbal and nominal speakers, it could never appear in the middle of the other constituents, and so the nominal speakers would never be tipped off that these constructions too were to be reinterpreted as verbal. In fact, one way to establish unequivocally that a given form in a Philippine or Formosan language is a noun (at least in some of its occurrences) is to find it used as a ligature attribute after another noun.

The remainder of this section will be devoted to a discussion of the reconstruction of the original functions of the individual 'focus' affixes and their development as verbal focus markers.

*-en

We reconstruct *-en as the ancestor of the 'Object Focus' or 'Goal Focus' suffix in Philippine languages. As with *ni-/in-, we assume that the primary function of *-en in Proto-Austronesian was to derive nouns from verbs and other nouns. For both deverbal and denominal nouns, the semantic effect of *-en derivation was 'future effect'. It is possible that both *ni- and *-en had begun to function as markers of verbal aspect in PAN, but if so, they had not become complementary allomorphs of 'Object Focus' in the way that their descendants now have in languages such as Tagalog.

Clear reflexes of this affix have not been identified in Oceanic, although PPN *kakano 'flesh, meat, pith' has been offered as a possible derivative of *kan 'eat' plus *-en. Reflexes of *-en in its nominalising function are however common in Western Austronesian languages of the Extra-Formosan group as well as in the Formosan languages.

The following examples from Ilokano show a neat contrast between reflexes of *-en and *ni-/in- in approximately the same derivational functions we reconstruct for PAN:

dengdeng-en	<i>'ingredients to be used in making a vegetable dish; that which is to become a vegetable dish'</i>
d-in-engdeng	<i>'the completed vegetable dish; that which has become a vegetable dish'</i>

***ni-/-in-**

The affix **ni-/-in-* functioned in Proto-Austronesian to derive nouns from verbs and other nouns, although it may have also begun to have the function of marking perfective aspect in verbs, a function which is now its primary one in Philippine languages. Based on evidence from Philippine and Formosan languages, both the prefix **ni-* and the infix **-in-* must be reconstructed for the earliest stage, with **ni-* infixing after all initial consonants except **l* and possibly **r*. The development of **-in-* as an infix seems to have preceded the development of infix **-um-* from **mu-*, judging from the reconstructible order **-umin-* (despite Wolff's **-inum-*; Pawley and Reid 1979:107).

***-ana**

The widespread *-an* and *-ana* suffixes marking locative nouns and verbs in Austronesian languages are reconstructed here as **-ana* rather than **-an*, primarily based on evidence from Oceanic languages and on the *-ana* suffixes found in Malagasy (Dahl 1976:118) and Tsou (Tung 1964:174-175). In the Central Pacific subgroup **-an(a)* is replaced by *-aŋa* (sometimes *-ŋa*). The substitution of the velar nasal for **n* is irregular but as this substitution has evidently occurred in several morphemes, it is very probable that the Central Pacific suffix is cognate with POC **-an(a)*. A similar correspondence is exhibited in certain languages of the eastern Solomons, which show *-aŋa* for expected *-ana*.

The original *-ana* was bimorphemic, probably consisting of the ligature *a* plus an attributive NP consisting of the demonstrative pronoun *na*. Several modern languages retain *-ana* in this function, including Bilaan. Compare Kagayanen *-an* < **-ana*, Ivatan *-ay* < *-a + ya*, and Isinai *-ad* < *?-a + di*. The latter are still demonstrative pronouns or definite articles, and illustrate the kind of process involved.

The combination of a linker plus a noun in PAN, as in modern languages, could only be attached to nouns, and this is reflected in the widespread appearance of reflexes of **-ana* as nominalising suffixes on verbal stems. Note that this implies that the bases for this affixation must have already been (\emptyset -derived) nouns when **a + *na* fused with them as suffixes. As we have noted elsewhere in this paper, this nominal status is the original and most widespread function of *-an*, with true verbal focus present only in a limited number of constructions with 'focus' affixes in Philippine languages.

***mu-/-um-**

The 'Agentive Focus' marker in Philippine languages is normally the infix *-um-*, a form which also occurs in some Malagasy AF verbs (Dahl 1973:118). We reconstruct the progenitor of this marker as **mu-/-um-*, based on data from

Formosan and Philippine languages which allow us to reconstruct both the infix and prefixed forms for Proto-Austronesian.

The original function of *mu-/-um- in Proto-Austronesian was probably that of deriving agentive nominalisations from nouns or verbs, a function very similar to that of -er in English.

The fact that m- and -m- forms in Atayal refer to animates (or atmospheric phenomena) supports this idea, since the agentive -er forms in English of course have the same implication. The nouns formed by this derivational process in PAN were later reanalysed as verbs by the process described elsewhere in this paper.

By the reanalysis process described below, reflexes of *mu-/-um- frequently derive transitive verbs in modern languages. That is, nominal structures such as 'John is the shooter of the bear' are reinterpreted as 'John shot the bear', and even originally intransitive verbs can become transitive via this route.

In Tsou, the *mu-/-um- prefix played a major role in deriving the active-passive distinction, and every Tsou sentence appears in one of these two modes, marked by m-prefixed auxiliaries and often m-verbs for active sentences, and m-less auxiliaries and verbs in passive structures.

The transition from *mu-/-um- as a nominaliser to *mu-/-um- as a verbaliser is analogous to the development of the other verbal foci, that is, it involved the reanalysis of the nominalised equational as a verbal construction.

The *-a verb suffix was one of the mechanisms used in Proto-Austronesian to derive verbs which 'focussed' on oblique actants from normal unmarked ergative verbs. It is directly reflected in the dependent ('subjunctive') OF suffix -a in Atayal (Egerod 1966:347) and in Tsou (Tung 1964:186). Dahl (1973:120) notes the use of -a in Malagasy as an imperative-optative AF affix, and states that -a is found in many languages with optative or subjunctive meaning, and also reconstructs it for PAN.

As in the case of the other foci, then, the verbalisation of mu- involves the substitution of the mu- form for the -a form in main clauses in all the daughter languages, and later on in subordinate clauses as well in many subgroups. The signal to the younger generation that the older generation had made the transition would be the occurrence of sentences with non-final subjects, constructions that are possible with a multibranching verbal structure but not with a binary equational nominal one. Thus the first Tagalog sentence below is ambiguously either nominal or verbal, but the second can only be verbal:

Bumili ng bigas ang babae
1 2 3 4 5

Bumili ang babae ng bigas
1 4 5 2 3

'The woman bought some rice'
4 5 1 2 3

*iSi- accessory focus

The prefix ?i- is a marker of Instrument, Benefactive, or Comitative Focus, and of Object Focus for 'transported objects' for a broad range of Philippine languages. We use the term 'Accessory Focus' as a convenient label for this

cluster of roles, though we retain the common abbreviation 'IF' ('Instrumental Focus') to avoid confusion with AF 'Agent Focus'. Based on evidence from Formosan languages, Dahl (1973:119) reconstructs this form as *Si- for PAN, in spite of the fact that this would be expected to produce hi- in Tagalog, rather than the ?i- that is actually attested.

The Formosan evidence for this reconstruction does not seem to be particularly clear and convincing. Dahl cites Amis IF sa- as one justification for the initial *S, but it turns out that Amis sa- (Chen 1982) is not a regular IF marker in Amis. Instead, Amis sa- derives instrumental nominalisations which only rarely occur in a construction which could be analysed as having an Instrumental subject. The implement-deriving sa in Rukai (Li 1973:274) would probably be a more tenable example. In Bunun, there is a similar form, but it is is- rather than si-, and marks future AF as well as IF. Assuming that the final vowel in this prefix was *i, rather than *a, the reconstruction of *iSi- provides a better explanation of the reflexes in Bunun and Philippine languages than does *Si-. Bunun is- can be accounted for as a result of vowel loss rather than metathesis, whereas Philippine ?i- forms can be assumed to have developed by reduction of the Philippine reflex *ihi- to *?i-. Northern Philippine languages which reflect PAN *S as glottal stop (or zero) would have reduced *i?i- to *?i-. A few Philippine languages still show hi- rather than ?i- as the IF prefix. Zorc (1977:134) cites Samar-Leyte, Waray, and Northern Samar hi- as forming part of the IF potential affix forms (nahi-, mahi-, etc.), and Tausug hipag- as the IF dependent, durative form.

We believe that the original function of *iSi- in Proto-Austronesian was, as in the case of the other reconstructed 'focus' affixes, nominalisation. In modern Philippine languages, it seems to be these Accessory Focus constructions that preserve the character of nominalised equationals even more so than other focus constructions.

10. Paradigm formation

The new denominal verbs formed by the reinterpretation of cleft sentence structures formed themselves into paradigms according to syntactic and semantic complementarity. One of the dimensions chosen in this regrouping was main clauses versus subordinate clauses. Thus in languages ancestral to languages such as Atayal (Egerod 1966) and Samareño (Wolff 1973), the new verbs were specialised to main-clause use, and the original *-i and *-a counterparts were confined to embedded environments. In Toba Batak (Van der Tuuk 1971), the -i/-aken reflexes were used in active sentences and the ni-/-en/-ana/-aken types were used as passives:

	Active	Passive
OF	-∅	ni- -in-
LF	-i	-an
IF	-hon	-hon

If subsequent investigation should reveal that the Toba Batak 'passives' are better analysed as nominalisations, and if the unmarked OF is really ergative, the Toba Batak system would be a quite close approximation to the one we have posited for the proto-language.

In most Northern and Central Philippine languages, the *-i and *-a forms were replaced completely by reflexes of the original deverbal nominalisers, and

the paradigms were composed according to aspect, with reflexes of the perfective *ni- infixes as *-in- to the Agentive, Locative, and Instrumental focus forms in *mu-/-um-, *-an, and *?i- respectively to produce perfective focus forms. For the Object Focus forms, *ni- and *-en derivatives were already OF and in complementary distribution with respect to aspect, so *ni-verbs assumed the perfective slot in the paradigm, with *-en forms filling the corresponding complementary non-perfective slots. This accounts for the unusual complementarity within the Tagalog OF paradigm between -in- infixes perfective forms and -(h)in- (< *-en) suffixed non-perfectives, without any necessity for an unmotivated morphological deletion transformation. A similar complementation process produced a different result in Kapampangan, where the ?i-prefixed OF form took over the future slot, the -in- form the perfective, and a reduplicated form the present progressive, resulting in the following paradigm:

i-sulat	'will write'
su-sulat	'are writing'
s/in/ulat	'wrote'

11. Development of verbal focus as a criterion for subgrouping

Austronesian languages can be characterised by whether or not their ancestors carried out this reinterpretation, and if so, how far they carried it. Tsou, for example, has no trace of a verbal focus system using originally nominalising affixes, though the nominal affixes are there in their original function. Languages such as Atayal and Samareño descend from systems in which the original verbal affixed forms were specialised to subordinate clauses, with the new verbs taking over main-clause focussing functions. Languages like Amis descend from languages which kept the *-i form only in the imperative and replaced all the others, and standard Tagalog replaced all the *-i type verbal forms by originally nominal affixes, while at the same time keeping the original deverbal nouns as well, resulting in a tremendous amount of structurally homonymous constructions that continue to confound us linguists to this very day.

The reinterpretation and replacement process was certainly a post-PAN innovation, but unfortunately the occurrence of this process by itself is unlikely to be very useful for subgrouping purposes, simply because once the stage was set, it became highly probable that the change would happen, and it could easily have happened independently in different languages. In the case of Amis, for example, it has only just begun to operate in a very limited set of environments, while in related Paiwanic languages, it seems to be well along. Similarly, the replacement of all *-i type forms was total in Standard Tagalog, but some of the original forms are still retained in Marinduque Tagalog, and this could hardly be taken as evidence that these different dialects belong to different higher-order subgroups. However, though the occurrence of the reinterpretation has very dubious subgrouping implications, it is still quite possible that specific idiosyncratic details may prove useful in this respect.

NOTE

1. This paper is a summary of some of the major concepts contained in a monograph which the authors are presently preparing. Because of the time constraints on a conference paper, the evidence for much of what we say here could not be presented. The extensive data from many areas of the family from which our conclusions are drawn will appear in the published monograph.

BIBLIOGRAPHY

ANCEAUX, J.C.

- 1952 *The Wolio language: outline of grammatical description and texts.*
The Hague: Nijhoff.

ASAI, Erin

- 1953 *The Sedik language of Formosa.* Kanazawa: Cercle linguistique de Kanazawa, Kanazawa University.

BLUST, Robert A.

- 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 9/2:1-15.

CAPELL, Arthur

- 1964 Verbal systems in Philippine languages. *Philippine Journal of Sciences* 93/2:231-248.

CENA, R.M.

- 1977 Patient primacy in Tagalog. Paper presented to the Annual Meeting, Linguistic Society of America, Chicago.

CHEN, Teresa M.

- 1982 Verbal construction and verbal classification in Nataoran-Amis. Ph.D. dissertation, University of Hawaii, in preparation.

CLARK, Ross

- 1973 Transitivity and case in Eastern Oceanic languages. In *Papers of the First International Conference on Comparative Austronesian Linguistics, 1974 - Oceanic.* *Oceanic Linguistics* 12:559-605.

DAHL, Otto Christian

- 1973 *Proto-Austronesian*. Scandinavian Institute of Asian Studies monograph 15. Lund: Studentlitteratur.

DUBOIS, Carl

- 1976 *Sarangani Manobo: introductory guide*. *Philippine Journal of Linguistics* special monograph 6. Manila: Linguistic Society of the Philippines.

EGEROD, Søren

- 1965 Verb inflection in Atayal. *Lingua* 15:251-282.
 1966 Word order and word classes in Atayal. *Language* 42:346-369.

ELKINS, Richard E.

- 1971 Western Bukidnon Manobo sentence structure. *Lingua* 27:216-262.

FERRELL, Raleigh

- 1974 Ergative, passive, or focus-marking? Elicitation of verb syntax in Indonesian languages. In: *Les langues sans tradition écrite: méthodes d'enquête et de description* [Actes du Colloque international du Centre National de la Recherche Scientifique], Nice, 28 juin-2 juillet 1971, 241-250. Paris: SELAF.

FILLMORE, Charles J.

- 1977 The case for case reopened. In: Peter Cole and Jerrold M. Sadock, eds *Syntax and semantics*, vol.8: *Grammatical relations*, 59-81. New York: Academic Press.

FOLEY, William A.

- 1976 *Comparative syntax in Austronesian*. Ph.D. dissertation, University of California, Berkeley.

GUZMAN, Videia P. De

- 1978 *Syntactic derivation of Tagalog verbs*. *Oceanic Linguistics* special publication 16. Honolulu: University Press of Hawaii.
 1979 Morphological evidence for primacy of patient as subject in Tagalog. Paper presented to the Annual Meeting, Linguistic Society of America, Los Angeles.

HARMON, Carol

- 1977 Kagayanen and the Manobo subgroup of Philippine languages. Ph.D. dissertation, University of Hawaii at Manoa.

HOU, John Yien-yao

- 1979 *Grammatical relations in Chinese*. Ph.D. dissertation, University of Southern California.

IKRANAGARA, Kay

- 1980 *Melayu Betawi grammar*. [Ph.D. dissertation, University of Hawaii at Manoa, 1975.] Jakarta: NUSA.

LI, Paul Jen-kuei

- 1973 *Rukai structure*. Academia Sinica special publications no.64.
Taipei: Academia Sinica.

LOPEZ, Cecilio

- 1941 *A manual of the Philippine national language*. 3rd edition. Manila:
Bureau of Printing.

MacDONALD, R. Ross and Soenjono DARDJOWIDJOJO

- 1967 *A student's reference grammar of modern formal Indonesian*. Washington,
D.C.: Georgetown University Press.

McFARLAND, Curtis D.

- 1978 Definite objects and subject selection in Philippine languages.
Studies in Philippine Linguistics 2/1:139-182.

McKAUGHAN, Howard P.

- 1958 *The inflection and syntax of Maranao verbs*. Publications of the
Institute of National Language. Manila: Bureau of Printing.

- 1962 Overt relation markers in Maranao. *Language* 38/1:47-51.

MIRIKITANI, Leatrice T.

- 1972 *Kapampangan syntax*. *Oceanic Linguistics* special publication no.10.
Honolulu: University Press of Hawaii.

MOREY, Virginia

- 1964 Distributional restrictions on co-occurrence of aspect and focus
morphemes in Ata verbs. *Oceanic Linguistics* 3/1:69-86.

MULDER, Jean and Arthur SCHWARTZ

- 1981 On the subject of advancements in the Philippine languages. *Studies
in Language* 5/2:227-268.

NAYLOR, Paz Buenaventura, ed.

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

PAWLEY, Andrew K., and Lawrence A. REID

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

REID, Lawrence A.

- 1979 Evidence for Proto-Philippine nominative marking. *Philippine Journal
of Linguistics* 10:1-20.

- 1981 Proto-Austronesian genitive determiners. In: Andrew Gonzalez and
David Thomas, eds *Linguistics across continents: studies in honor of
Richard S. Pittman*, 97-105. Manila: Summer Institute of Linguistics
and Linguistic Society of the Philippines.

- 1982 The demise of Proto-Philippines. See pp.201-216 in this volume.

SCHACHTER, Paul and Fe T. Otnes

1972 *Tagalog reference grammar*. Berkeley: University of California Press.

STAROSTA, Stanley

1974 Causative verbs in Formosan languages. In: *Papers of the First International Conference on Comparative Austronesian Linguistics, 1974 - Proto-Austronesian and Western Austronesian*. *Oceanic Linguistics* 13:279-369.

1979 The end of Phrase Structure as we know it. *WPLUH* 11/1:59-76.

TOPPING, Donald M.

1973 *Chamorro reference grammar*. PALI language texts, Micronesia. Honolulu: University Press of Hawaii.

TUNG, T'ung-ho

1964 *A descriptive study of the Tsou language, Formosa*. Institute of History and Philology special publication no.48. Taipei: Academia Sinica.

TUUK, H.N. van der

1971 *A grammar of Toba Batak*. The Hague: Martinus Nijhoff. (Translation by Jeune Scott-Kemball of *Tobasche Spraakkunst*, Amsterdam 1864 and 1867.)

WOLFF, John U.

1973 Verbal inflection in Proto-Austronesian. In: Andrew B. Gonzalez, ed. *Parangal kay Cecilio Lopez (Essays in honor of Cecilio Lopez on his seventy-fifth birthday)*, 71-94. *Philippine Journal of Linguistics* special monograph issue no.4. Quezon City: Linguistic Society of the Philippines.

1979 Verbal morphology and verbal sentences in Proto-Austronesian. In: Naylor, ed. 1979:153-168.

ZORC, R. David Paul

1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. PL, C-44.

ATAYALIC FINAL VOICED STOPS

Paul Jen-kuei Li

1. INTRODUCTION

The Atayalic group of Formosan languages comprises Atayal and Sediq, each with dialects of its own.¹ Atayal, in turn, consists of two major dialects, Squliq and C'uli'. Since Squliq dialects are all fairly uniform, whereas C'uli' dialects are divergent from each other, only one Squliq dialect, Pyasan, will be represented but several C'uli' dialects, Skikun, Mnawyan, Paljawan, Maspazi', Mayrinax, Matabalay and Sakuxan will be included in this study. The Sediq dialects, Inago, Taroko, Toda, Tojan and Truwan will be represented. I have collected in the field extensive lexical materials and some data for verb inflections for the above dialects in this report.

Most of the previous publications on Atayal, such as Ogawa (1931), Egerod (1965a,b, 1966), Yamada and Liao (1974) and Ferrell (1979), were based on Squliq, the most innovated dialect in the sense that it has lost a lot of information for historical reconstruction. For one thing, quite a few word-initial consonants and most unstressed vowels are lost; see Li (1980a). For another, some consonants including the voiced stops, i.e. /b, r, z, g/, may not occur word finally; see Egerod (1966:122).

However, Mayrinax, a C'uli' dialect recently investigated, retains /b, g/ (phonetically [β, γ] respectively) in the word-final position, e.g. /ginabugab/ 'rice gruel', /sabsab/ 'eaves', /ʔabag/ 'leaf', /basag/ 'millet', /qag/ 'ribs'.

2. PROTO-ATAYALIC PHONOLOGY

Before dealing with the specific problems of Proto-Atayalic *-b, *-d, *-g' and *-g, let us examine a reconstruction of Proto-Atayalic phonology briefly. Given below are the sound correspondences of five or six Atayal dialects and five Sediq dialects.

Distribution of a certain correspondence is indicated by the hyphen, e.g. /p-/ indicates that it occurs prevocally or preconsonantly, i.e. word initially or word medially; /-p/ indicates that it occurs postvocally, i.e. only word finally; /k/ indicates that it occurs in all positions. Relevant conditions or phonetic environments for a certain correspondence may be specified under each correspondence.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 171-185. *Pacific Linguistics*, C-75, 1982.

© Paul Jen-kuei Li

171

PAN		*p		$\begin{cases} *t \\ *t(T) \end{cases}$			*k	*q	
PA		*p-	*-p	*t-	*t-	*-t	*k	*q	*?
Ata	Sq	p-	-p	t-	c-	-t	k	q	?
	Ms	p-	-p	t-	t-	-t	k	?	?
	Sk, Mn	p-	-p	t-	c-	-c	k	q	?
	Mx	p-	-p	t-	t-	-t	k	q	?
	Pl	p-	-k	t-	t-	-c	k	(?)	?
Sed	Tn	p-	-k	t-	t-	-c	k	q	?
	Td	p-	-k	t-	t-	-c	k	q	?
	Tr	p-	-k	t-	c-	-c	k	2	?
	Tk	p-	-k	t-	c-	-c	k	q	?
	In	p-	-k	t-	c-	-c	k	q	?

Conditions: /__i

PAN		*b		$\begin{cases} *d \\ *d(D) \\ *d'(Z) \end{cases}$			*γ (R)						
PA		*b	*-b	*d-	*d-	*-d	*g-	*-g-	*-g-	*-g	*-g	*-g	*-g
Ata	Sq	b-	-p	r-	r-	-?	g-	-g-	∅	-y	-w	-w	-w
	Ms	b-	-p	r-	r-	-t	g-	-g-	(g)	-y	-w	-w	-w
	Sk, Mn	b-	-p	r-	r-	-?	g-	-g-	∅	-y	-x	-x	-x
	Mx	b-	-b	r-	r-	-?	g-	-g-	(g)	-y	-g	-g	-g
	Mt	b-	-p	r-	r-	-t	g-	-g-	∅	-g	-g	-g	-g
Sed	Pl	b-	-k	ř-	ř-	-c	g-	-g-	∅	-y	-w	-w	-w
	Tn	b-	-k	d-	d-	-c	g-	-r-	-r-	-y	-w	-∅	-∅
	Td	b-	-k	d-	d-	-c	w-	-r-	-r-	-?	-?	-w	-w
	Tr	b-	-k	d-	j-	-c	g-	-r-	-r-	-g	-g	-w	-w
	Tk	b-	-k	d-	j-	-c	g-	-r-	-r-	-g	-g	-w	-w
In	b-	-k	d-	j-	-c	g-	-r-	-r-	-g	-g	-g	-g	

Conditions: /__i /a_a /a_u /i_# /u_# /a_#

PAN	*g'(j)			*c(C)	*s(S)	{*h *t'}	*l	*t(N), *ñ		
PA	*-g'-	*-g'-	*-g'	*c	*s	*x	*h	*r	*r	*l
Ata Sq	-r-	-r-	-y	s-	s	x	h	y	z	l
Ms	-s-	-s-	-y	s-	s	x	h	y	z	l
Sk	-s-	-s-	-s	c-	s	x	h	y	y	l
Mx	-s-	-s-	-y	c-	s	x	h	w	w	l
Mt	-s-	-s-	-g							
Pl	-ř-	-ř-	-y	c-	s	x	h	r	r	l
Sed Tn	-g-	-y-	-y	c-	s	x	h	r	r	l
Td	-w-	-y-	-?	c-	s	x	h	r	r	l
Tr	-g-	-y-	-g	s-	s	x	h	r	r	l
Tk			-g	s-	s	x	h	r	r	l
In	-g-	-y-	-g	s-	s	x	h	r	r	l
Conditions:	/i__a							/__i		

PAN	*n	*m		*ŋ	*w	*y
PA	*n	*m-	*-m	*ŋ	*w	*y
Ata Sq	n	m-	-m	ŋ	w	y
Ms	n	m-	-m	ŋ	w	y
Sk	n	m-	-m	ŋ	w	y
Mx	n	m-	-m	ŋ	w	y
Pl	n	m-	-ŋ	ŋ	w	y
Sed Tn	n	m-	-ŋ	ŋ	w	y
Td	n	m-	-ŋ	ŋ	w	y
Tr	n	m-	-ŋ	ŋ	w	y
Tk	n	m-	-ŋ	ŋ	w	y
In	n	m-	-ŋ	ŋ	w	y

PAN		*a	*i	*u	*ə	* <u>-aw</u>		* <u>-ay</u>	
PA		*a	*i	*u	*ə	*aw	*-aw	*ay	*-ay
Ata	Sq	a	i	u	∅	o	-aw	e	-ay
	Ms	a	i	u	a	aw	-aw	ay	-ay
	Sk	a	i	u	∅	o	-aw	e	-ay
	Mx	a	i	u	∅	aw	-aw		-ay
	Pl	a	i	u	∅	o	-aw	e	-iy
Sed	Tn	a	i	u	e	o	-o	e	-e
	Td	a	i	u	ə	o	-aw	e	-ay
	Tr	a	i	u	ə	o	-aw	e	-ay
	Tk	a	i	u	ə	o	-aw	e	-ay
	In	a	i	u	ə	o	-ag	e	-ay
Conditions:						/non-final syllables		/non-final syllables	

3. PROTO-ATAYALIC *-b

By comparative study, it is clear that Proto-Atayalic (PA) *-b has been devoiced as /-p/ or /-k/² in all the dialects except Mayrinax in the Atayalic group, for example³

1. Proto-Atayal *h-um-gub > Mx /h-um-gub/, Sq /h-m-gup/, Sk /h-m-op/, Ms /h-am-agup/, Mn /h-m-up/, Pl /h-um-uk/ 'to do magic'
2. Proto-Atayal *ma-gaub > Mx /g-um-aub/, Sq /m-gop/, Sk, Mn /m-gop/, Ms /ma-gaup/, Pl /ma-gagok/ 'to share one cup while drinking'
3. Proto-Atayal *h-um-ab > Mx /h-um-ab/, Sq /h-m-op/,⁴ Sk, Mn /h-m-ap/ 'to stab'
4. Proto-Atayal *pə-sihub > Mx /p-sihub/, Sq /shup/, Sk /p-shup/, Ms /pa-siyup/ 'to suck'

These are the verbs that ended with *-b in Proto-Atayal, though it might not have gone so far back as Proto-Atayalic. The main problem in reconstructing these forms for Proto-Atayalic is that the cognates do not appear in both Atayal and Sediq subgroups.

Mayrinax /-b/ is rare. I have found only six items ending with /-b/ among over 1,000 items I have collected, and only four of them have cognates, as listed above.

In the dialects where *-b has been devoiced as -p or -k, the verbs exhibit b ~ p or b ~ k alternations:

5. Sq /h-m-gup, hbg-an/,⁵ Sk /h-m-op, hub-an/, Pl /h-um-uk, hub-an/, Mt, Sx /h-um-agup, hagab-an/ 'to do magic'
6. Sq /m-gop, gob-an/, Sk /m-gop, gob-an/, Ms /ma-gaup, ?in-gaub-an, gaub-i/, Pl /ma-gaguk, pa-gagub-an/, Mt, Sx /ma-gaup, pa-gaub-an/ 'to share one cup'
7. Sq /h-m-op, hab-an/, Sk /h-m-ap, hab-an/ 'to stab'
8. Sq /shup, shub-an/, Sk /p-shup, p-shub-an/ 'to suck'

In these alternations, /b/ always appears in the word-medial position immediately followed by a suffix, whereas /p/ or /k/ appears word-finally. They all share the same rule, synchronically and diachronically:

$$9. b \rightarrow \left\{ \begin{array}{l} p \\ k \end{array} \right\} / _ \#$$

The following verb also exhibits the same $b \sim p$ or $b \sim k$ alternation in these dialects:

- 10a. Sq, Sk and Mn /q-m-atap, qtab-an/, Pl /ʔ-um-atak, ʔatap-i/⁶ 'to cut with scissors'
 b. Ms /k-am-arip, karib-an/, Sx /k-um-arip, k-in-rib-an/,
 Tk /q-m-arik, qrib-i/⁷ 'to cut with scissors'

Although the Mayrinax form /paqway/ 'to cut with scissors' is not a cognate with the forms in these dialects, the Proto-Atayalic form for the verb should have ended with *-b as indicated by both the Atayal and Sediq dialects. Two variant proto-forms for the verb stem, *qatab and *karib, seem to have existed.

Three verbs exhibit $b \sim k$ in Sediq dialects alone:

11. Proto-Sediq *r-um-bərub > Td /r-m-bərək, rbrəb-i/, Tn /r-m-berək, rbrub-i/, Tr /r-m-bərək, bərb-i/ 'to broil'
 12. Proto-Sediq *mələb > Td /məluk, ləb-i/, Tn /meluk, leb-i/ 'to close'
 cf. PHN *q₂əNəb as reconstructed by Tsuchida (1976:168).⁸
 13. Proto-Sediq *s-um-arub > Td /s-m-aruk, srub-i/ 'to burn hair'

The verb 'to yawn' is a cognate shared by most dialects in the entire group:

14. *ma-surab > Sq, Sk, Mn /m-suyap, syab-i/, Ms /ʔam-suyap/, Mt, Sx /ma-suyap/, Mx /ma-suwag/, Tn /ma-surak/, Tn /m-suwak/,⁹ Td, Tr, Tk /m-surak/ 'to yawn'

This verb also exhibits $b \sim p$ alternation in some Atayal dialects. Unfortunately the Mayrinax form /masuwag/ ends with /-g/ instead of /-b/. The form needs some explanation: The semivowel /w/ may have been derived from *r and the final consonant /-g/ is probably the result of dissimilation from the labial *b, viz. Proto-Atayalic *masurab > *masuwab > masuwag.

4. PROTO-ATAYALIC *-g

The derivations of PA *-g in the various dialects are conditioned by the preceding vowel. After *u, PA *-g is retained as /-g/ in two Atayal dialects (Mayrinax and Matabalay) and three Sediq dialects (Inago, Truwan and Taroko), devoiced as /-x/ in Skikun and Mnawyan, become /-ʔ/ in Toda, and changed to /-w/ in most other dialects:

15. *lubug > Mx, Mt, Tr, Tk, In /lubug/, Sk, Mn /lubux/, Td /lubuʔ/, Sq, Ms, Tn /lubuw/ 'Jew's harp'
 16. *sinirug > Mx, Mt /siniyug/, Tr, Tk /snirug/, Sk, Mn /sniyux/, Sq /sniyuw/, Tn /sniruw/ 'necklace, string'

After *a, PA *-g is retained as /-g/ only in Mayrinax, Matabalay and Inago, devoiced as /-x/ in Skikun and Mnawyan, and changed to /-w/ in most other dialects:

17. *siyag > Mx, Mt, In /siyag/, Sk, Mn /syax/, Tn /siyoi/, Sq /syaw/, Ms, Td, Tr, Tk /siyaw/ 'side, shore'

After *i, PA *-g is retained as /-g/ only in Matabalay, Truwan Taroko and Inago; it has become /-y/ in most other dialects:

18. *bunaqig > Mt /buna[?]ig/, Tk, In /bnaqig/, Sq, Sk, Mn /bnaqiy/,
Mx /bunaqiy/, Td /bnaqi[?]/, Tn /bnaquy/, Tr /bna[?]ig/, Sx, Pl /buna[?]iy/
'sand'

In the dialects in which word-final /-g/ may not occur, they all show the morphophonemic alternations $g \sim w$ or $g \sim x$ in some of the verbs:

19. In Squliq:

active	passive imperative	
hmaw	hag-i	'to scoop out'
muhuw	lhug-i	'to thread a needle'

20. In Maspazi[?]:

rahaw	rahag-i	'to trap'
mayuw	yug-i	'to exchange'

21. In Skikun and Mnawyan:

malhax	tlhag-i	'to wake'
mnbux	nbug-i	'to drink'

22. In Sakuxan:

ramahaw	rahag-i	'to sharpen (a knife)'
[?] amasuw	[?] asug-i	'to distribute'

23. In Palḡawan:

gumuṛaw	guṛag-i	'to choose'
lumpuw	lapag-i	'to count'

24. In Truwan:

dmayaw	dyag-i	'to help'
smapaw	spag-i	'to spread (a mat)'

25. In Taroko:

dmayaw	jyag-i	'to help'
gmaaw	gæag-i	'to choose'

In these alternations, /g/ occurs in the word-medial position followed by a suffix, whereas /w/ or /x/ occurs word finally. They can be stated by the rule:

26. $g \rightarrow \begin{Bmatrix} w \\ x \end{Bmatrix} / __\#$

In short, there is sufficient internal evidence in these dialects for *-g to occur in the word-final position at an earlier stage.

A note in passing: both Truwan and Taroko retain /-g/ after vowels /i/ and /u/, so the alternation $g \sim w$ is found only after vowel /a/.

5. PROTO-ATAYALIC *-g' (*-z OR *ɣ)

However, not all /-g/'s in modern dialects go back to an early *-g.¹⁰ Some /-g/'s have derived from a different source, PA *g' (or reconstructible as *ɣ or *z):

27. *ma-barig' > Mt /mabazig/, Tr, Tk /marig/, Sk, Mn /mbes/, Sq /maziyl/, Ms /mabaziyl/, Mx /mabaay/, Tn /maruy/, Td /mari?/ 'to buy'
28. *kægig' or *kærig' > Mt /kagig/, Tr, Tk /kærig/, Sk, Mn /kgis/, Sq, Mx /kgiy/, Ms, Pl /kagiy/, Tn /keguy/, Td /kæri?/ 'hemp'

The two examples above show a different set of correspondence from that of *-g discussed before. The difference in the word-final position lies mainly in Skikun and Mnawyan /-s/ vs. /-x/ : /-s/ < *-g', /-x/ < *-g. The distinction is lost in all the other dialects except Mayrinax, which shows a contrast between /-y/ < *-g' and /-g/ < *-g after certain vowels. The above two cognates and their derivatives appear in the whole group. More cognates can be identified in the Atayal dialects:

	'to air'	'to escape'	'leggings'	'to lie on one's side'
Sq /-y/	mahiy	magyay	sragiy	mtkkiy
Ms /-y/	ramahiy	magiyay	sragiy	(matkakiil)
Sk /-s/	rmahis	mgyas	cragis	mtkkis
Mx /-y/	rumahiy	magiyay	--	(masikakil)
Mt /-g/	rumahig	(magiyay)	--	(matakakil)
Pl /-y/	řumahiy	magiy	--	maskakiy

No example has been found in the word-initial position.

In the word-medial position quite a few cognates appear in the whole group, and with two different sound correspondences from that in the word-final: one between i_a¹¹ and another elsewhere:

		'how many'	'a little later'	'reed of loom'	'yesterday' ¹²	'Buy!'	'Peel hemp!'
Ata	Sq /-r-/	pira?	kira?	bzira?	hira?	bir-i	kgir-i
	Ms /-s-/	pisa?	kisa?	bagisa?	su-hisa?	bais-ani	kamgis-i
	Sk /-s-/	pisa?	kisa?	bgisa?	c-hisa?	bes-i	kgis-i
	Mx /-s-/	(piya?)	kisa?	bagisa?	cu-hisa?	baynay	kumkagis-i
	Mt /-s-/	pisa?	kisa?	bagisa?	su-hisa?	binas-i	kamkagis-i
	Pl /-ř-/	piřa?	kiřa?	bagiřa?	hiřa?	binař-i	kakagiř-i
Sed	Tn /-g-, -y-/	piya?	kiya?	bgiya?	ciga?	brig-i	
	Td /-w-, -y-/	piya?	kiya?	bwiya?	chiwa?	briw-i	
	Tr /-g-, -y-/	piya?	kiya?	bgiya?	s-higa?		
	In /-g-, -y-/	piya?	kiya?	bgiya?	sigā?	brig-i	
PA	*-g'-	*pig'a?	*kig'a?	*bagig'a?	*cu-hig'a?	*barig'-i	*kumakagig'-i

Note that PA *-g'- is less well preserved in the environment between i_a (only well preserved in the item 'yesterday') than in the other environments in the Sediq dialects.

The three sets of correspondences discussed in this section are all derived from the same source historically. The verb inflections in the various dialects also indicate that they are etymologically equivalent:

29. The alternation $r \sim y$ in Squliq:

stem	active	passives	
baziy	maziy	b-n-ir-an	'to buy'
kgiy	mkgiy	kgir-un, k-in-gir-an	'to peel hemp'
bahiy	mahiy	h-n-ir-an	'to air'
pgyay	mgyay	pgyar-an	'to escape'

30. The alternation $\check{r} \sim y$ in Palḡawan:

biniy	miniy	bina \check{r} -i	'to buy'
pagiy	magiy	pa \check{r} er-i	'to escape'
ṙahiy	ṙumahiy	ṙahir-an	'to air'

31. The alternation $s \sim y$ in Maspazi²:

baziy	mabaziy	bais-ani	'to buy'
kagiy	kmkagiy	kingis-an	'to peel hemp'
rahiy	ramahiy	kimmagis-an	'to air'

32. The alternation $s \sim y$ in Mayrinax:

baay	mabaay	binas-un	'to buy'
kgiy	kumakgiy	kumkagis-i	'to peel hemp'
rahiy	rumahiy	rahis-an	'to air'

33. The alternation $s \sim g$ in Matabalay:

bazig	mabazig	mabinas-an	'to buy'
kagig	kumakagig	kamkagis-i	'to peel hemp'
	rumahig	rahis-ani	'to air'

34. The alternation $w \sim \text{?}$ in Toda:

bari [?]	mari [?]	briw-i, briw-an	'to buy'
-------------------	-------------------	-----------------	----------

In the alternations of these dialects, /r, ṙ, s, w/ occur in the word-medial position, whereas /y, g, ?/ occur in the word-final position. The changes in the dialects can be stated by the rules:

35. $r \rightarrow y$
 36. $\check{r} \rightarrow y$
 37. $s \rightarrow y$
 38. $s \rightarrow g$
 39. $w \rightarrow \text{?}$
-

To reconstruct Proto-Atayalic *g', it is worth noting the fact that /s/ is phonetically an alveopalatal fricative [š] in all the modern Atayal dialects. In addition to /s/, the semivowel /y/ also reflects the proto-phoneme as a palatal. Proto-Atayalic *g' has merged with *g in most Sediq dialects and in one Atayal dialect, Matabalay. An unusual change is the devoicing of *g' to /-s-/ word medially, yet the voicing is still preserved word finally in Matabalay.

Both Dempwolff and Dahl reconstructed Proto-Austronesian (PAN) *g' for the phoneme under discussion. The derivations PAN *g' > PA *g' > g in some modern Atayalic dialects seem quite straightforward.

6. PROTO-ATAYALIC *-d

Although /-d/ is not retained in any modern dialect, a few Sediq verbs exhibit $d \sim c$ or j (palatalised before /i/) $\sim c$ alternation; actually only the following verbs have been found to exhibit the alternation:

	active	active imperative	passive imperative	passive ₁	passive ₂	
40. Torjan:	hmaŋuc	haŋuc	hŋed-i			'to cook'
Toda:	hmaŋuc	haŋuc	hŋəd-i	hŋəd-un	hnŋəd-an	
Taroko:	hmaŋuc	haŋəc	hŋəj-i			
Truwan:	hmaŋəc	haŋəc	hənj-i ¹³	hənd-un	hənd-an	
Inago:	hmaŋəc	haŋəc	hŋəj-i			
41. Torjan:	rmutuc	rutuc	rtud-i			'to join, link'
Toda:	rmutuc	rutuc	rtud-i	rtud-un	rtud-an	
Taroko:	mrutuc	rutuc	ptuj-i			
Truwan:	rmutuc	rutuc	rtuj-i	rtud-un		
Inago:	rmutuc	rutuc	rtuj-i	rtud-un		
42. Truwan:	dməruc	dəruc	drəj-i	drəd-un	drəd-an	'to rub, whet'
Inago:	dndoruc	doruc	drəj-i	drəd-un		

These alternations can be stated by the rules:

R(a). $d \rightarrow j / __ i$

(b). $d \rightarrow c / __ \#$

Parallel to the other two voiced stops /b, g/, the dialects that no longer have these stops in the word-final position do show $b \sim p$ (or k) and $g \sim w$ (or y) alternations, so it can be inferred that /d/ also must have occurred in the word-final position at an earlier stage, at least in Proto-Sediq since Sediq dialects show d (or j) $\sim c$ alternation.

Atayal has no /d/. Sediq /d/ corresponds to Atayal /r/ preconsonantly and prevocally except before /i/. Before /i/ (or /y/), *d has become /j/ in Inago, Truwan and Taroko dialects of Sediq.

		'pygmy deer'	'horse'	'beard'	'hornet'	'earwax'
Ata	Sq /r/	para?	rme?	ŋurus	tryuŋ	qpuri?
	Ms /r/	para?	ramai?	ŋurus	tariyuŋ	?apuri?
	Sk /r/	para?	m-rme?	ŋurus	tryuŋ	qpuri?
	Mx /r/	para?	ramai?	ŋa-ŋurus	--	qapuri?
Sed	Tn /d/	pada?	dmai?	ŋudus	tɕiyuŋ	qpudi?
	Td /d/	pada?	dmai?	ŋudus	tɕiyuŋ	qpudi?
	Tr /d,j/	pada?	dmai?	ŋudus	tɕiyuŋ	?puji?
	Tk /d,j/	pada?	--	ŋudus	tɕiyuŋ	qpudi?
	In /d,j/	pada?	dmai?	ŋudus	tɕiyuŋ	qpudi?
PA	*d	*pada?	*damai?	*ŋudus	*tɕiyuŋ	*qapudi?

However, Atayal /r/ corresponding to Sediq /d/ does not occur word finally. In that position, neither /d/ in any Sediq dialect nor /r/ in many Atayal dialects may occur.

Based on internal reconstruction, we have noted that Sediq /d/ (or /j/ before /i/ in some dialects) regularly changes to /-c/ word finally; see also Yang (1976:649-650). However, Sediq /-c/ has three possible origins, *-t, *-d

or *-c. Let us first examine the correspondence for *-t in the word-final position:

		'fang'	'side dish'	'flying squirrel'	'eight'
Ata	Sq /-t/	waqit	ramat	yapit	spat
	Sk /-c/	waqic	ramac	yapic	spac
	Ms /-t/	waʔit	ramat	yapit	ma-sapat
	Mx /-t/	waqit	ra-ramat	apit	ma-ma-spat
Sed	Tn /-c/	waqic	damac	rapic	ma-sepac
	Td /-c/	waqic	damac	rapic	ma-səpac
	Tr /-c/	waʔic	damac	rapic	ma-səpac
	Tk /-c/	waqic	damac	rapic	ma-səpac
	In /-c/	waqic	damac	rapic	ma-səpac
PA	*-t	*waqit	*damat	*rapit	*ma-səpat

In these examples *-t has become /-c/ in most dialects but is retained as /-t/ in a few Atayal dialects, among which Mayrinax is the dialect that usually preserves the older features. Those dialects that end with /-c/ all show t ~ c alternation and share the same rule below:

43. t → c / ___#

For example

44. In Skikun:

miŋac	ŋat-an	'to rob'
kmac	kat-an	'to bite'

45. In Toŋan:

qmiyuc	qyut-i	'to bite'
rmeŋac	rŋat-i	'to growl'

46. In Toda and Inago:

qmiyuc	qyut-an	'to bite'
smaqic	saqit-an	'to cut with scissors'

Now let us examine another set of correspondence with /-c/ in all Sediq dialects, but with /-t/ or /-ʔ/¹⁴ in Atayal dialects:

		'wing'	'mouse'	'leopard'	'ashes'	'mulberry'
Ata	Sq /-ʔ/	paliʔ	qoliʔ	kliʔ	qbuliʔ	tliuʔ
	Sk, Mn /-ʔ/	paliʔ	qoliʔ	kliʔ	qbuliʔ	tluʔ
	Ms /-t/	--	qawlit	yakalit	ʔabulit	--
	Mx /-ʔ/	paliʔ	qawlit	ʔakliʔ	qabuliʔ	--
	Mt, Sx /-t/	pali-huy	ʔawlit	yakalit	ʔabulit	taliuʔ (Sx)
Sed	Tn /-c/	palic	qolic	rkelic	qbulic	(tdiyuc)
	Td /-c/	palic	qolic	rkəlic	qbulic	(tdiyuq)
	Tr /-c/	palic	ʔolic	rkəlic	ʔbulic	(kdiyuc)
	Tk /-c/	--	qolic	rkəlic	qbulic	(kdiyuc)
	In /-c/	palic	qawlic	raklic	qbulic	(kdiyuc)
PA	*-d(?)					

In the above correspondence, only three Atayal dialects regularly end with /-t/, most other Atayal dialects end with /-ʔ/, and all the Sediq dialects end with

/-c/. Are the final consonants /-t/, /-ʔ/ and /-c/ in modern dialects reflexes of an earlier *-d or *-c? The former is more likely because it seems more natural for a voiced stop to become a voiceless stop (*-d > -t) in the word-final position, than for an affricate to change to a stop (*-c > -t). Also the change of the dental stop *-d > -t, -ʔ or -c parallels that of the labial stop *-b > -p or -k in devoicing.

Some supporting evidence for the analysis adopted in this paper can be drawn from the following derivations:

47. PAN *paʔid > PA *palid > Sediq /palic/, Atayal /paliʔ/, Yami /panid/, RukBd /parídi/, Itb, Ivatan, Masaka /panid/ 'wing'
48. Proto-Hesperonesian *[tT]aNiu[DZ]¹⁵ > PA *taliud > Atayal /tliuʔ/ 'mulberry' cf. RukBd /taliodo/

Since we have PA *-b and *-g as discussed in the preceding sections, the symmetric pattern of distribution will also allow us to reconstruct PA *-d for the correspondence ʔ-t-c as illustrated above. There is no reason to create a new phoneme for this particular correspondence. The residual problem is that we are still missing a correspondence for *-c in the PA system. However, there is no reason to expect all consonants to occur word finally.

7. SUMMARY AND ATAYAL IMPLICATIONS FOR PAN

The four final voiced stops *-b, *-d, *-g' and *-g reconstructible for Proto-Atayalic are reflexes of nearly the same Proto-Austronesian phonemes:

PAN	PA	Modern Atayalic
1. *-b >	*-b >	-b, -p, -k
2. *-d >	*-d >	-t, -ʔ, -c
3. *-g' >	*-g' >	-g, -y, -s, -ʔ
4. *-ɣ >	*-g >	-g, -w, -y, -x, -ʔ

The phonetic features of these four voiced stops are much better preserved in the word-initial and -medial position than in the word-final position in modern Atayalic dialects. In addition to the regular sound correspondences and symmetric pattern of distribution, the morphophonemic alternations: 1. b ~ p and b ~ k, 2. d ~ c, 3. r ~ y, ʔ ~ y, s ~ y, s ~ g and w ~ ʔ, 4. g ~ w and g ~ x of the verbs also provide important evidence for the occurrence of these voiced stops in the word-final position in Proto-Atayalic.

Proto-Atayalic *-b is rare, just as PAN *-b is rare. Dempwolff reconstructed only 15 PAN forms ending with *-b, and seven of them were reduplicated forms.¹⁶ Compare Archaic Chinese that has only a few forms with *-b and Proto-Indo-European that has few forms with *b in any position. Similarly, Proto-Atayalic *-g and *g' are also rare.

Atayal has some interesting implications for PAN reconstruction. What has hitherto been reconstructed as PAN *-aw by Dyen (*-av by Dempwolff) shows up as /-ag/ in Inago, a Sediq dialect in the Atayalic group, e.g. PAN *laŋaw > raŋag 'fly'. The reconstruction of *-aw for PAN may have to be revised as ending with some type of consonant phonetically similar to /-g/. Of course, there is always the possibility that the Inago /-g/ was an analogical innovation in the dialect by changing all forms ending with /-w/ to /-g/.¹⁷ However, only the older speakers preserve /-g/ in their speech, whereas the younger speakers tend

to change it all to /-w/. The ongoing change in Inago is just like the other Atayalic dialects in which a similar change has taken place involving /-g/ derived from Proto-Atayalic *-g, which in turn derived from what hitherto has been reconstructed as PAN *-ɣ.

NOTES

1. This paper was written with the support of the National Science Council Grant NSC-68H-03-01(06), Taiwan. I have greatly profited from discussion with Shigeru Tsuchida, who informed me of the existence of the important dialects included in this study. I am also indebted to Robert Blust, David Zorc, Pang-hsin Ting and Hwang-cherng Gong for their valuable comments and suggestions for improvement on an earlier draft of this paper.
2. Labial consonants may not occur word finally in any Sediq dialects and in two Atayal dialects, Palŋawan and Mstbaun, as well as in the speech of younger speakers of Skikun and Mnawyan. In these dialects, all labials become velars word finally; see Li (1980a) and Yang (1976).
3. The following is a list of abbreviations of languages and dialects used in this article:

Ata	Atayal	Mx	Mayrinax	Sx	Sakuxan
In	Inago	Pl	Palŋawan	Td	Toda
Itb	Itbayatan	RukBd	Rukai (Budai)	Tk	Taroko
Mn	Mnawyan	Sed	Sediq	Tn	Toŋan
Ms	Maspazi?	Sk	Skikun	Tr	Truwan
Mt	Matabalay	Sq	Squliq		

The morpheme boundary is indicated by the hyphen in the cited forms.

4. Squliq has the vowel alternation a ~ o in the verb /h-m-op, hab-un, hab-i, hab-aw/ 'to stab'; see Li 1980a:399.
5. The verb stem and suffix are separated by a hyphen. Note the metathesis of /b/ and /g/ in the Squliq forms /hmgup, hgup, hbg-an, hbg-un, hbg-i/ 'to do magic'.
6. The stem-final *b may have become /p/ in /ʔatap-i/ in Palŋawan by analogy.
7. The initial /q-/ in Taroko is irregular.
8. The Atayal dialects such as Squliq share the same cognate, though not the alternation: Sq /q-m-luʔ, qlʔ-un, qlʔ-i/ 'to close'. However, the Atayal dialects seem to have gone one step further than Sediq in the sound change: *-b > *-d > -ʔ; see section 6 for the Atayal reflexes of PA *-d. PHN stands for Proto-Hesperonesian in Tsuchida's system.

9. The Toŋan form /m-suwak/, which is in discrepancy with the forms in other Atayal dialects, can be regularly derived from PAN *ma-suwab. Perhaps PA *r in *ma-surab is the result of change by the male speaker. See Li (1980b) for a detailed account of the differences between male and female speech in the Mayrinax dialect.
10. I am indebted to Shigeru Tsuchida for pointing out to me the two different sets of correspondence and hence two different origins. Some of the cognates, especially 'a little later', 'women's basket', 'part of loom', and 'to lie on one's side', were identified by him first.
11. Perhaps the relevant phonetic environment is simply /__a; cf. PAN *pag'ay > In, Td, Tk, Tr (all Sediq) /payay/ 'rice plant'. However, Sq, Ms, Sk (all Atayal) /pagay/ does not correspond regularly to Sediq.
12. Note that /h/ is lost in the Sediq dialects of Toŋan and Inago, all in Nant'ou prefecture. However, it is still preserved in the Sediq dialects of Pribaw /c-higa/, Skadaŋ, Raus and Swasal /s-higa/, all in Hualien prefecture, as based on Tsuchida's (1980) data.
13. Note that the velar ŋasal /ŋ/ has been partially assimilated to the following dental as /n/.
14. Note that /-t/ in Mx /qawlit/ and /-ʔ/ in Sx /taliuʔ/ are irregular.
15. The reconstructed form is based on Tsuchida (1976).
16. Dempwolff reconstructed the following PAN forms ending with *-b: *dibdib 'to suck on the breast', *lənəb 'to clarify', *kəbkəb or *kubkub 'to cover', *kulub 'to heat', *ləbləb 'overflow', *l(u)kub 'protection', *tabtab 'to knock', *ta(ŋ)kub 'suited', *təyəb 'numerous', *t'ibt'ib or *t'ubt'ub 'slide off', *lunqab 'to open', *luḷub 'to burn', *luvab 'steam'.
17. As Robert Blust (personal communication) pointed out, the change *(-)w- > g occurs in other Austronesian languages, so why not *-w > g here? If we eliminate PAN *-aw we will be left with an asymmetric diphthong system:

uy	iw
(ey?)	(ew?)
ay	--

BIBLIOGRAPHY

BLUST, Robert A.

- 1973 Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages' - 2. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 5/3:33-61.

DAHL, Otto Chr.

- 1973 *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series 15. Lund: Studentlitteratur.

DEMPWOLFF, Otto

- 1938 *Vergleichende Lautlehre des austronesischen Wortschatzes* 3. *Zeitschrift für Eingeborenen-Sprachen*, Beiheft 19. Berlin: Reimer.

DYEN, Isidore

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

EGEROD, Søren

- 1965a Verb inflexion in Atayal. *Lingua* 15:251-282.
 1965b An English-Atayal vocabulary. *Acta Orientalia* 19:203-220.
 1966 A statement on Atayal phonology. *Artibus Asiae Supplementum* XXIII (Felicitation volume for the 75th birthday of Prof. G.H. Luce) 1:120-130.

FERRELL, Raleigh

- 1979 Phonological subgrouping of Formosan languages. In: Paz Buenaventura Naylor, ed. *Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages*, 241-254. Ann Arbor: The University of Michigan.

LI, Paul Jen-kuei

- 1980a The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 51/2:349-405.
 1980b Men's and women's speech in Mayrinax. In *Papers in honor of Prof. Lin Yü-k'eng on her seventieth birthday*, 9-17. Taipei: Wen Shin.
 1981 Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52/2:235-301.

OGAWA, Naoyoshi

- 1931 *Atayal vocabulary*. Taihoku: Gouvernement Général of Taiwan.

REID, Lawrence A.

- 1971 *Philippine minor languages: word lists and phonologies*. *Oceanic Linguistics* special publication 8. Honolulu: University of Hawaii Press.

TSUCHIDA, Shigeru

- 1976 *Reconstruction of Proto-Tsouic phonology*. *Study of Languages and Cultures of Asia & Africa, Monograph Series* 5. Tokyo.
- 1980 Linguistic position of Sikikun and Manawyan — linguistic bases of subgrouping Sqolyeq and Ts^ʔole dialects in Atayal. MS.

YAMADA, Yukihiro and Ying-chu LIAO

- 1974 A phonology of Tayal. *Research Reports of the Kochi University* 23/6:109-117.

YANG, Hsiu-fang

- 1976 The phonological structure of the Paran dialect of Sediq. *Bulletin of the Institute of History and Philology, Academia Sinica* 47/4:611-706.

PROTHESIS IN THE LANGUAGES OF CENTRAL MALUKU:
AN ARGUMENT FROM PROTO-AUSTRONESIAN GRAMMAR

James T. Collins

0. INTRODUCTION

Although Brandstetter and earlier scholars (including Van der Tuuk) considered in some detail the reconstruction of PAN grammar, Dempwolff did not take up the topic. For several decades it was laid aside while our conception of the PAN sound system was expanded and refined. Happily in recent years there has been a revival of interest in PAN syntax.¹ Furthermore the importance of "moribund morphology" (Hamp 1973) in surviving AN languages has become more apparent.² Given the expanding corpus of PAN grammatical markers and our evolving understanding of PAN syntax, it is time to further exploit that knowledge of ancient morphosyntactic systems in assessing sound changes. The paper presented here proposes a reanalysis of unexplained initial sounds in certain languages of Central Maluku (Indonesia). A careful analysis of available evidence suggests that these prothetic elements are retentions of the PAN noun markers, *si and *u.

1. THE PROBLEM

Stresemann's book, *Die Lauterscheinungen in den ambonischen Sprachen*, contains a section dealing with phonetic phenomena in word-initial positions (1927:114-118). Here he considers unexpected initial segments in non-verbal lexical items. He claims that these sounds occur, especially before initial a, as a development of the "Spiritus lenis". He also notes that Central Maluku languages are not unique in displaying this phenomenon. He cites similar developments in Gorontalo, Aiduma, Kei, Kor and Teor.

Chlenov and Sirk (1973:75-76) briefly discuss those cases where an unexpected initial h appears in certain Central Maluku languages. They say this h "most probably ... is not a reflex of some primary initial consonant but a prothetic sound which precedes the initial vowel". Further reference is made to Stresemann's "Spiritus lenis" argument. There the issue is closed.

It is unsatisfactory to appeal to hypothetical onsets in order to resolve residual problems as Stresemann did. It is no more acceptable to apply a label to a phenomenon and then dismiss the problem as if resolved. If possible,

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, 187-200. *Pacific Linguistics*, C-75, 1982.

© James T. Collins

attempts should be made to uncover the basis for a certain phenomenon; naming problems should not replace solving them. Chlenov and Sirk's solution by nomenclature and a reference to Stresemann should be reconsidered in light of the greater amount of relevant data now available.

Elsewhere (Collins 1980 and 1981) it has been demonstrated that in some cases the retention of a reflex of PAN *q in initial positions accounts for the occurrence of h- in Kamarian, Rumakai, Haruku, Saparua and Asilulu. A second major source of 'unexpected' initial segments is considered here, namely the retention of ancient noun markers. Two sets of correspondences are considered: (1) h-, y- and Ø and (2) w- and Ø. In each set the evidence points to grammatical markers retained under certain phonetic or sociosemantic conditions.

2. THE CORRESPONDENCE h-, y- AND Ø : PAN *si

From Saparua and Nusalaut, Chlenov and Sirk (1973:76) cite hahunno 'turtle-dove' which they assume displays prothetic h. This assumption is based on the vowel-initial entries in Asilulu and Hila, ahune. In fact, there are a few more words in Saparua similar to hahunno. These words compare with similar entries in Paulohi, Ruta and Sepa-Teluti. Here we cite data from the Latu dialect of Saparua and for the sake of comparison the relevant entries in another related language, Asilulu. Entries from Proto-Central Maluku (PCM) and PAN are also given.

PAN		Latu	Paulohi	Ruta	Sepa	Tehoru	Asilulu
*kabil	'hook'	hahillo	ahire	ahirolo	yahilo	yahilo	ahil
*kaba(Ct)	'thread'	haho	haha	ahailo	yaha	yaha	aha
*kəbkəb	'cover' ³	hohano	hohane	uhano	yohano	yohano	ohan
PCM							
**kaba	'oar' ⁴	ahē	hahate	aha	payano	hayano	ahat
**abun	'dove'	hahune	hahune	(mahano)	---	---	ahun
**abatū	'a bamboo'	hahatu	---	aha [?] uro	yahatu	yahatu	---

In general, the initial h in Latu and Paulohi corresponds to y- in Sepa and Tehoru; in Ruta and Asilulu we find zero.⁵ (The appearance of m- in the entry for 'dove' as well as the irregular vowel suggest that Ruta word is not a reflex of PCM **abun.) Latu and Paulohi display a single exception each: Latu ahe 'row' and Paulohi ahire 'hook'.⁶ In Sepa and Tehoru a distinctive word-specific metathesis took place in the entry for 'oar'. In addition Sepa displays a prenasalised reflex of *b. That is:

PCM **kaba > aba > yaba > baya > Sepa mbaya > paya-no
 Tehoru baya > haya-no

The PCM and PAN words are all of the form (k) [+voc] [-high] b [+voc] ([-voc]). At an earlier stage in the history of these languages *k was lost; so all the words began with a non-high vowel. Similarly in all these languages *b became h. Thus we find h, y or Ø before the sequence [+voc] [-high] h [+voc]. The overall similarity in the phonetic arrangement of each word as well as the striking regularity in each language's treatment of this sequence strongly suggest that some sound was present at an earlier stage. This sound yielded h in Latu and Paulohi, y in Sepa and Tehoru and zero in Asilulu and Ruta.

The conditions for this correspondence are quite specific. In other entries, however, Sepa and Tehoru y- does not correspond to h- in Latu and Paulohi.

PAN		Latu	Paulohi	Ruta	Sepa	Tehoru	Asilulu
*aŋin	'wind'	anino	anine	ani:no	yanino	yanno	anin
*aku	'1sg'	au	au	au	yau	yau	a'u
*kaSiw	'wood'	ai	ai	ai	yai	yai	ai
*kaSu	'2sg' ⁷	ale	ale	ale	yale	yale	ai
*kami	'1pl excl'	ami	ami	ami	yam	yam	ami
PCM							
**kambat	'sea-eel' ⁸	apato	---	apa'olo	yapato	yapato	apat
**ambat	'wound'	apatiro	lapate	iapa'iro	yapato	yapan	apat ⁹
**aqi	'leg' ¹⁰	ai	wai	ai	yai	yai-m	ai
**amet	'sago bark'	ameto	amete	ame'eolo	yameto	yameto	---
**aman	'village' ¹¹	amanno	amane	ama:no	yamano	yamano	---

With few exceptions, we note that Latu, Paulohi, Ruta and Asilulu display zero where Sepa and Tehoru have y-. Again all entries are nouns; all of them have initial a (after the loss of k- at an intermediate stage). By comparing the above table with the one on the preceding page, we can infer that the conditions for the correspondence of Latu h- and Sepa-Teluti y- were more specific. In Sepa-Teluti y- appears before nouns with word-initial non-high vowels, usually a.¹² In Latu and Paulohi h- appears only when the initial sound of a noun is a non-high vowel followed by h (from *b). Ruta and Asilulu display zero in all these cases.

To explain these correspondences shall we reconstruct a PAN sound? The fact that in a number of cases we know that PAN *k was the original initial sound rules out this approach. PAN, as it is reconstructed, did not allow word-initial consonant clusters. What other possible analysis can account for the observed regularity in sound correspondence?

The critical observation here is that these words are nouns. The fact that the unknown sound whose reflexes are y- and h- and zero could not have been present as a consonant cluster in PAN points to the affixation of some element to these nouns. The most likely addition to a noun is a determiner or article of some sort. In contemporary Central Maluku languages there are at least two noun-markers: PCM **-ə 'singular noun-marker' and PCM **-a 'plural noun-marker'. These are suffixed to nouns. In some languages these markers are optional or their use is rule governed. In other languages the noun-marker is no longer productive.¹³ Could there have been an earlier noun-marker which was prefixed and then later became non-productive just as some contemporary Maluku languages display suffixed, non-productive noun-markers?

Of all the Austronesian languages there is one family which is extraordinarily rich in preposed noun-markers. Every Philippine language has perhaps a half-dozen such markers. Indeed, Reid (1978) proposed eleven Proto-Philippine determiners, distinguished for nominative, genitive and locative cases and common and personal nouns. Among the recurring determiners is si.

In Maranao, Cebuano, Inibaloi and Pangasinan, for example, si appears as the topic and nominative personal marker (Reid 1978:51). In Tagalog Schachter (1972:76) identifies si as the personal noun-marker used with proper names, kinship terms and titles. Of the thirty-two languages listed in Reid (1978:39) all but nine display si as the 'personal subject determiner'. It is no wonder that Reid reconstructs *si as the Proto-Philippine nominative marker.

This Proto-Philippine *si and all other such determiners have an important interrelationship with the complex voice systems of Philippine languages. These

complicated voice systems or focus systems are found in the languages of the Philippines and in languages in nearby island groups (Guam, Formosa, Borneo and Sulawesi) as well as Madagascar. Elsewhere there is little evidence of a productive focus system in contemporary AN languages. Nonetheless there is some evidence which indicates the existence of focus systems at earlier periods in other AN languages.

Our knowledge of PAN syntax is still at a very rudimentary stage. But even if we set aside the question of the antiquity of the focus system we are still confronted with these interesting noun-markers and determiners because they appear in AN languages which no longer possess an active focus system. One noteworthy non-Philippine language which displays noun-markers which correspond to Philippine noun-markers is Malay.

In Malay there are at least two noun-markers: *saŋ* and *si*. Wilkinson calls *saŋ* "a titular prefix" and *si* "a demonstrative prefix". In modern Malay *si* is often used contemptuously in derogatory reference. Wilkinson observes that "this contemptuous use of *si* is neither ancient nor universal". He notes the retention of *si* in the neutral pronouns *si-anu* 'what's his name' and *si-apa* 'who'; he adds that *si* "is sometimes used very deprecatingly of oneself ... or one's young son".

In Sumatran dialects of Malay *si* never acquired the pejorative connotation it has in most peninsular Malay usages. Rather, "it appears in names of the great spirit *si-kulambai*, of the great dragon *si-katimuna*, of the famous buffalo *si-bĕnuang sakti* and of the unconquered cock *si-kunani*. It is commonly used in personal names and titles". This older usage is confirmed by perusal of old spells, invocations and mantras found in Skeat (1900:590-598) and elsewhere.

Kamus Dewan adds other interesting uses of *si* in peninsular Malay. In addition to the derogatory sense mentioned above, it is used in nicknames and nursery names. When used in reference to oneself it has the force of self-denigration. It is also used to stress the agent or patient of the sentence. It occurs in verbalised nominalisations.¹⁴ This dictionary also cites a number of plant and animal names in which *si* is a non-productive prefix: *sidiŋin*, *sikĕdidi*, *sitawar*.

In Indonesian the use of *si* is more widespread and carries no deprecatory force. Echols and Shadily say *si* is "a k.o. definite article used before names (as form of reference); before non-proper nouns and before words which describe s.o. or s.t. (especially used as pet names)". I have further observed its use in colloquial Indonesian with singular pronouns, for example *si-dia* (article + 3sg) 'that certain someone'.

Other languages in western Indonesia also display the use of *si* with nouns and pronouns. In Batak, Bali and Madura (where it is *se*) the usage is very similar to Malay.¹⁵ In Batak it also appears frequently in nominalisations (Van der Tuuk 1971:74-75) and personifications such as in toponyms and zoological nomenclature (Warneck 1977:226). In numerous languages of south-east Sulawesi **si* seems to be prefixed to pronouns.¹⁶ In Kayan (Borneo) it appears before personal names and kinship names.¹⁷

There is considerable evidence, then, which justifies the reconstruction of **si* 'personal noun-marker' in the proto-language of the western languages of Austronesia. It is argued here that this **si* was partially retained in several languages of Central Maluku. In these languages **si* was not restricted to personal names or pronouns.¹⁸ The conditions of its usage are not clear at this point. Nevertheless the wide range of its occurrence in Malay and other

languages indicates that it may never have been restricted exclusively to personal nouns. In fact, there are some suggestions that it may have been a respectful or personalising noun-marker. At any rate, it is proposed here that *si was retained in the proto-language of Central Maluku and some of its descendants.

In the ancestral language of Asilulu this *si was lost. But in Sepa-Teluti it was retained as y-. In Paulohi and Saparua it appears as h- but only in the specified environment discussed above. The limited appearance of y- in Sepa-Teluti and h- in Saparua and Paulohi requires some explanation.

First of all we note that PAN *s is regularly retained in all these languages. Here we claim that *si- became h- or y-. The fact that *si- was a morphological marker made it possible for *si- to undergo innovations which did not affect *s in other cognates. Special treatment of morphological affixes is well known in other language families. But, in addition to that general consideration, here it appears that *si- was subject to a different phonetic environment in that it occurred in the pretonic position.

Some support for this analysis comes from a word in Sepa-Teluti. Latu displays *sahuto* 'fog, mist'; Haya also displays *sahuto*. In Tehoru both *sahuto* and *yahuto* occur; in Sepa only *yahu* occurs. Either we reconstruct a PCM form ***sabut* 'mist' and assume that in Sepa-Teluti this *s was interpreted as an occurrence of *si- or we presume that these are all reflexes of PAN *kabut 'fog'. If we accept the latter interpretation then Sepa displays the regular shift of *si to y- before a. The occurrence of *si as s in Latu is unexplained. Perhaps complex borrowing is involved here. No matter what the ultimate source of ***sabut*, the alternation of *yahuto* and *sahuto* indicates that *si could have become y- in Sepa-Teluti.

The retention of the PAN noun-marker *si in Sepa-Teluti, Latu (and other Saparua dialects) and Paulohi was conditioned by the phonetic environment. In words in which *si- was preposed to words with initial *a or *ə the resulting vowel sequence of [+high] [-high], led to its retention. Elsewhere it was lost.

That is *si- > hi-/# [-high]
> Ø-/# [+high]

In Latu and Paulohi ***hi-* was retained only when the first consonant of the following word was h (from *b); so, for example, ***hi+ah* became *hah*. In Sepa-Teluti ***hi* became ***i-* and later y- : ***hi-* > ***i-* > y-.

3. THE CORRESPONDENCE w- AND Ø : PAN *u

Having explored the likelihood that h-, y- and Ø are reflexes of the PAN noun-marker *si, there remain several other words which display unexpected initial segments. These suggest another PAN noun-marker, *u.

PAN		Latu	Paulohi ¹⁹	Ruta	Sepa	Tehoru	Asilulu
*asu	'dog'	asuro	asu	asu:lo	wasu	wasu	asu
*aka	'elder sibling'	wa:	wa?a	wao:	wa?	wa?an	a?a-
*qa(z)ay	'jaw'	wala	'ala'	wala	walan	alan	
PCM							
**asu	'cheek' ²⁰	wasu	'fasu'	asu	wasu	(fafan)	(ka?ika)
**katan	'tongs'	atanno	atane	a?ano	watano	watano	atan

In Sepa-Teluti there are two prothetic segments: y- and w-. It has been argued that y- is a reflex of *si-. If that is so, what is the reflex of w-? Setting aside the possibility of some additional PAN sound, there are two other possible analyses. One would argue that *si became y- or w- depending on the phonetic environment. The other would suggest that w- is the reflex of another PAN noun-marker. Like the retention of *si, the retention of this noun-marker was conditioned by the phonetic environment.

In considering the first analysis it is important to note that at an early stage, PAN *Z became **d. Elsewhere (Collins 1980) I have argued that **d was retained in Sepa-Teluti until a late stage. Therefore in Sepa-Teluti *qa(z)ay had become **qa^hda. With that in mind, it is clear that w- appears before words containing anterior stops or fricatives: wasu, watano and **wa^hda (later walan). The only known exception is the reflex of *aka 'elder sibling'. This irregularity in the kinship system (often marked by irregularities in other language families) is probably caused by contamination from other kinship terms. The parallelism with wali 'younger sibling (of the same sex)' and wato 'spouse of parent's sibling of the opposite sex' is striking.

So in the first analysis *si became w- when it preceded a low vowel followed by a coronal occlusive; *si became y- when it preceded a low vowel followed by segments other than coronal occlusives. Hence, the overriding factor which triggered the retention of *si and its subsequent sound changes was dissimilation.

Another analysis should be considered. This too involves dissimilation. Among Reid's reconstructed noun determiners for Proto-Philippines we note *ʔu, the common nominative determiner (Reid 1978:53). Elsewhere he also notes the existence of the noun-markers ʔu in Ami, uʔ in Seediq and o in Tsou, all Formosan languages (Reid 1979:30-31). We might assume that *u can be reconstructed in PAN or at least in some language ancestral to Central Maluku, Formosan and Philippine languages.

We conclude that *u was retained in Sepa-Teluti as w-. Here too dissimilation was involved. The high vowel *u was retained before non-high vowels; likewise the back vowel *u was retained near anterior (coronal) consonants. That is [^{+VOC}_{+high}] was retained before the sequence [^{+VOC}_{-high}] [^{-VOC}_{+COR}]; elsewhere it was lost.

Bearing in mind the numerous noun markers which have been reconstructed in Proto-Philippines, the occurrence of two noun-markers in the proto-language of Central Maluku is not surprising. That their retention might be conditioned by dissimilation is also reasonable. Furthermore by arguing for the presence of the noun-marker *u in PCM we can approach other difficult problems.

In addition to the body parts wala 'jaw' and wasu 'cheek', Latu also displays worujjo 'gall'. Chlenov and Sirk (1973:74) have argued for "the derivation of the initial w in these words from a hypothetical prefixal u or wa- perhaps originating from PAN *buaq 'fruit' as a kind of counter".²¹ Under closer scrutiny it appears that w- in the Latu reflex of *qapəju could be a reflex of *u 'noun-marker'. It was retained under the circumstances specified earlier: a non-high vowel followed by a coronal occlusive or fricative; thus **u + əju > wəju > woru-jjo. If that is so, Latu has apparently retained *u only in the semantic category of body parts.²² As in other AN languages the noun-marker was retained only in certain restricted vocabulary items.

This interpretation sheds some light on some problems in other closely related languages. Some words in these languages display unexpected initial w-. Note the following entries from three languages of westernmost Seram.

PAN		Asilulu	Manipa	Wakasihu
*qapəju	'gall'	welu	wailu	ilona
*pəru	'tortoise'	wenu	wenu	wino
*qiSu	'shark'	weu	weu	wiwo
*(qS)a(rR)u	'spirit', ²³	walu	---	walua

The entries from Asilulu require no comment. Every sound change is regular except for the lowering of *i between two high vowels in the entry for 'shark' (that is *u+qiSu > uqiu > uiu > ueu > weu). In Manipa the entry for 'gall' is striking. Elsewhere (Collins 1980) I have argued for the late retention of a reflex of *q in Manipa (and other languages of Central Maluku). In the reflex of *qapəju, then, the sequence of events must have been as follows: *u+qapəju > uqapəju > uapəju > wapəju > waelu > wailu. In other words the a of wailu is a retention of the *a of *qa. The following e from *ə was raised to i through dissimilation (after a). The entry from Wakasihu displays unexplained loss of w-. That it may have been present at an earlier stage is suggested by the shift of **e to i, perhaps through dissimilation as in Manipa. However, the shift of **e to i in the following two entries points to assimilative raising due to the high vowel in the second syllable. (The late suffixation of **-ə 'noun-marker' and later regular sound changes (Collins 1980) account for the final o in wino and wiwo.)

Apparently the phonetic and semantic conditions for retaining *u in these three languages differed from those in the Central Maluku languages further to the east. In Sepa-Teluti the conditions are strictly phonetic. In Latu only the nouns of the semantic category of body parts which fulfill certain phonetic requirements display retention of *u. In view of the four entries from Manipa, Wakasihu and Asilulu, it seems that different phonetic and semantic categories are involved. PAN *u was retained when it occurred before words which contained a vowel followed by a continuant and the round vowel *u.²⁴ At least three of the entries suggest that the retention occurred only before nouns of a certain category. Both the shark and the tortoise have sacral importance in local traditions.²⁵ It may be that *u was retained only as a respectful noun-marker.²⁶

4. CONCLUSION

Tentatively we conclude that the ancestral language of Central Maluku retained *si and *u, noun-markers. The conditions for their retention in each language have been suggested here. This analysis is not without its problems. Nonetheless in these pages an attempt has been made to resolve the persistent problem of accretion of unexpected segments in initial positions of PAN words.²⁷ This is a phenomenon which also appears in numerous Oceanic languages and even some western Austronesian languages. Further refinement of the argument regarding Central Maluku languages is necessary but the solution offered here represents a significant advance over other approaches.

Dyen's solution (1962) to a similar problem in Niala, an unidentified language in western Seram,²⁸ is unacceptable. He noted that Niala displayed *asu > wasu 'dog' and *anak > wana 'child'. Based on these data and their partial correspondence to Chamorro data, Dyen's conclusion was that there existed a previously undiscerned PAN sound *W.

Dyen and Stresemann used similar tactics in approaching the problem of unexpected initial segments in Central Maluku languages. Each observed the

problem and attributed unexpected sounds to a dubious proto-sound, the one "tentative" *W, the other "Spiritus lenis". The solution proposed here is a less radical one. Some of the unexpected initial segments noted by Stresemann appear to be retentions of PAN noun determiners *si and *u. Both these grammatical markers already have been reconstructed in the proto-language ancestral to the languages of western Austronesia, as well as Formosan languages. Thus, instead of insisting on the reconstruction of new sounds in epicyclical fashion, we have attempted to explore the possible role of these existing PAN noun-markers in the descendants of PCM.

The less dramatic analysis attempted here yields results of varying strength which in turn require still more analysis and testing. Still the implications for PAN are perhaps more far-reaching than the reconstruction of yet another poorly attested proto-sound. If it is true that the retention of PAN noun-markers was the source of otherwise unexpected sounds in these Central Maluku languages, then we have uncovered a new and possibly productive starting point for considering similar phenomena in other Austronesian languages.²⁹ In which case, not only are troublesome sound changes likely to be resolved but we move a step closer to the reconstruction of PAN syntax, a syntax which includes a range of noun determiners.

ACKNOWLEDGEMENTS

My thanks go to Bob Blust who first drew my attention to apparently similar phenomena in Oceanic languages, to Gérard Diffloth and Eric Hamp who discussed earlier drafts with me and to Laurie Reid who commented on the portion of my dissertation which considered this problem.

NOTES

1. See in particular Blust 1974, Foley 1976, Pawley and Reid 1976 and Pawley 1977.
2. Refer to Bradshaw 1978, Geraghty 1978 and Blust 1979.
3. In PCM, PAN *kəbkəb became **kəb+an 'thwart seat'. Note the permanent affixation of the locative affix -an which should be compared to similar fossilised suffixes in other AN languages; for example in Aru we find entries for 'roof cover' *buŋbuŋ > ꞑuꞑuŋan (Ujir) and ꞑuꞑunan (Ngaibor). The semantic shift from 'cover' to 'thwart seat (i.e. which covers the stern of a canoe)' is not extreme.
4. Tentatively this entry is reconstructed with an initial **k based on the occurrence of kaha 'oar' in Buru; this is probably a loanword.
5. In Ruta (Amahai) this prothetic h- was perhaps lost at a late date. Stresemann (1927) recorded Amahai hoha 'a kind of pandanus' while I recorded in Tehoru yoha 'pandanus with a long leaf'. This suggests an earlier form *si+oba.
6. The noun in Latu [aʔaheto] suggests that in this language the morphological process which creates instrumental nominalisations by reduplicating the initial sound of the verbal root may take precedence over the prothetic h-. Note for example tuo 'sit' and tutuoto 'kitchen stool'; naʔi 'place under' and nanaʔito 'wedge'; sosa 'rub vigorously' and sososato 'brush'.
7. The forms cited here probably reflect the early affixation of the demonstrative **le; so, *kaSu > kau > ka > ʔa > a+le > ale.
8. This is reconstructed with initial **k based on the entries in another PCM descendant, Soboyo; we note several sea-eel species under compound nouns headed by kabac. Further confirming evidence is found in Banda Eli (Kei): kambat 'sea-eel'.
9. In contemporary Asilulu the meaning is restricted; it refers only to wounds suffered as a result of circumcision.
10. Proto-Eastern Oceanic has been reconstructed with *waqe 'foot, leg' (Pawley 1972:42).
11. It may be that this word is derived from PAN *ama 'father'.
12. In section three another initial segment (w) is discussed.
13. For example, Luhu (South-westernmost Seram) retains noun-markers under certain circumstances.
14. For example, we note ber+si+bisu (middle+article+'mute') 'to act as if mute'. Perhaps this should be compared to the Toraja use of si. When it is preposed to a verb, it marks reciprocity: sitiro 'see each other'; siboboʔ 'fight each other' when prefixed to numerals it means 'each'; for example, daʔdua 'two' beside sidaʔdua 'two each'.

15. In Javanese and Old Javanese (Kawi) *si* appears as an "article used for persons" (Zurbuchen 1976:124).
16. In my fieldnotes of Popalia (Tukangbesi Islands) *i-* occurs unexpectedly in some pronouns. Zainuddin Untung (1979) records similar "excrecent" *y-* in Bonerate, a south-east Sulawesi language spoken in islands near Salayar. Note: Popalia *iaku* '1sg'; *ikoʔo* '2sg'; and Bonerate *yaku* '1sg'; *yikoʔo* '2sg'; *yikita* '1pl incl'; and *yikami* '1pl excl'.
17. Ding Ngo notes a paradigm (*deklinasi*) in which the genitive is marked by *n-* and the vocative by zero; this applies to personal names and kinship terms (Lii' Long 1972:1xxiv). S. Morgan (personal communication) suggests that *h-* is the reference marker. Note the following examples:

Personal names:

Kinship terms:

<i>aŋin</i>	<i>haŋin</i>	<i>naŋin</i>	<i>ine</i>	<i>hine</i>	<i>nine</i>	' <i>mother</i> '
<i>naan</i>	<i>hnaan</i>	<i>nnaan</i>	---	<i>hareʔ</i>	<i>nareʔ</i>	'title (female)'
<i>ukaŋ</i>	<i>hukaŋ</i>	<i>nukaŋ</i>	<i>ibaan</i>	<i>hibaan</i>	<i>nibaan</i>	'title (male)'
<i>ului</i>	<i>hului</i>	<i>nului</i>				

In these cases *h-* seems to derive from **s*; we observe **salaq* > *halaaʔ* 'incorrect' and **sumpit* > *hmput* 'blowpipe'.

18. In Sepa and Tehoru no noun appears with initial *a-*. Either *y-* or *w-* occurs where other Central Maluku languages display zero. We conclude, then that a noun-marker occurred before all nouns but was retained only before *a-* (and possibly ***-ə*). It is difficult to determine whether only certain semantic classes were involved. A glance at the wordlists presented here suggests that pronouns, body parts, flora and fauna, important tools and natural substances may have been classified in a special respect-avoidance category.
19. Words in quotation marks are taken from Stresemann (1918). The few surviving speakers of Paulohi (who number less than twenty) have replaced these words with Malay loanwords. Ninety per cent of Paulohi speakers died in the earthquake and tidal wave of 1898.
20. Paulohi points to ***pasu* but Sepa-Teluti suggests ***asu*. The Paulohi form may be a compound from ***upan* 'face' and ***asu* 'cheek'. However, Blust (personal communication) notes several forms in other AN languages which suggest **pasu*.
21. This recalls Codrington's (1885:71-72) argument regarding similar phenomena in Nengone and Gaua (and perhaps Marshallese). Nonetheless even here we note in Gaua "a demonstrative article" *u* (1885:368).
22. The occurrence of *w-* in Latu *ni waru-n* 'his spirit' probably belongs here as a body part in the broadest sense. The Latu entry for 'shark' *woku-ro* requires comment in another paper.
23. This reconstruction is based on the Central Maluku words and the Malay words *aru* 'evil spirit assuming human shape' and *haru* 'plaguing ... especially of spirit'. In Central Maluku languages *walu/waru* usually refers to the spirit of the recent dead, especially one's relatives.
24. It is possible that the PCM reflex of PAN **qiSu* was, in fact, ***qəSu* (with unexplained and distinctive centralisation of the first vowel). If that is so, then, just as in Sepa-Teluti and Latu the initial vowel must have been a non-high vowel (*ə* or *a*).

25. Some clans revere the shark as an ancestor. Members of these clans are forbidden to eat its flesh. The shark is, of course, an agent of retribution; in numerous stories the sea tortoise appears as an agent of rescue at sea. In the recount of a recent shipwreck (recorded in Serua, South-east Maluku) the sea tortoise seems to have represented a dead parent. In connection with the role of shark and tortoises it is important to recall that *walu/waru* is usually the spirit of a recently deceased relative.
26. Whether the controlling factor was semantic ([+respect] vs. [-respect]) or grammatical (agentive vs. objective) is not yet clear.
27. As Blust has pointed out (1979:214-215) earlier Austronesianists did not hesitate to propose "old articles" as the source of unexpected initial segments. That these proposals were unproven has been well argued by Blust. Rather than propose hypothetical particles otherwise not reconstructed, this paper points to familiar PAN noun-markers.
28. Through the generosity of W.A.L. Stokhof, in 1978 I was able to see a partial copy of the 1937 Niala wordlist which apparently was Dyen's source. At the time, I concluded that Niala was clearly *not* a language of western Seram. Since then, the volume containing the complete Niala vocabulary has been published (Stokhof 1981). The classification of this language is still a problem. The informant, Matheus Soeplatoe, was a native speaker of Alune; he was the head of a village near Riring, the military garrison and prison of West Seram in the colonial period. But the language is not a language of Seram. Based on my notes of the Nila language, spoken on a small volcanic island of that name, south of the Banda islands, 'Niala' is a slightly distorted version of Nila. How Bpk Matheus learned Nila remains a puzzle. He may have served as a soldier or catechist there or he may have learned the language through a prisoner in the forced labour camp in Riring. A quick perusal of my Nila data suggests that [β] occurs before u and o (either from PAN *u or resulting from loss of *p). If this is so, its occurrence before *a in Nila may be related to the phenomena discussed in this paper. Of course, considerably more research in the languages of south-west Maluku is necessary to test that hypothesis.
29. It is striking that Zorc (1981) independently reached similar conclusions regarding *W in Formosan languages.

BIBLIOGRAPHY

ASIS SAFIOEDIN

- 1977 *Kamus Bahasa Madura-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

BALAI PENELITIAN BAHASA SINGARAJA

- 1975 *Kamus Indonesia-Bali*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

BLUST, Robert A.

- 1974 Proto-Austronesian syntax: the first step. *Oceanic Linguistics* 13:1-15.
- 1979 Proto-Western-Malayo-Polynesian vocatives. *Bijdragen tot de Taal-, Land- en Volkenkund* 135:205-251.

BRADSHAW, Joel

- 1978 Tonogenesis in Jabem. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 10/1:125-140.

BRANDSTETTER, Renward

- 1916 *An introduction to Indonesian linguistics*. Four essays translated by C.O. Blagden. London: Royal Asiatic Society.

CHLENOV, M.A. and Ülo SIRK

- 1973 Merger of labial phonemes in Ambonese languages. *Acta et commentationes Universitatis Tartuensis, Oriental studies* 2/1:58-97.

CODRINGTON, R.H.

- 1885 *The Melanesian languages*. Oxford: University Press.

COLLINS, James T.

- 1980 The historical relationships of the languages of Central Maluku, Indonesia. Doctoral dissertation, University of Chicago. To appear in *PL*, series B.
- 1981 Preliminary notes on Proto-West Central Maluku: Buru, Sula, Taliabo and Ambelau. In: R.A. Blust, ed. *Historical linguistics in Indonesia*, part 1. *NUSA* 10:31-45.

DEMPWOLFF, Otto

- 1934-38 *Vergleichende Lautlehre des austronesischen Wortschatzes*. *Zeitschrift für Eingeborenen-Sprachen*, Supplements 15, 17, 19. Berlin: Reimer.

DYEN, Isidore

- 1962 Some new Proto-Malayopolynesian initial phonemes. *JAOS* 82:214-215.

ECHOLS, John M. and Hassan SHADILY

- 1970 *An Indonesian-English dictionary*. Jakarta: Bhratara.

FOLEY, William A.

- 1976 Comparative syntax in Austronesian. Doctoral Dissertation, University of California, Berkeley.

GERAGHTY, Paul

- 1978 Topics in Fijian language history. Doctoral dissertation, University of Hawaii.

HAMP, Eric

- 1973 Prenasalization in Eastern Oceanic. *Oceanic Linguistics* 12:295-301.

ISKANDAR, T.

- 1970 *Kamus Dewan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

LII' LONG

- 1972 Syair Lawe' bagian pertama. Manuscript, edited and translated by A.J. Ding Ngo, Mendalam-Putussibau.

PAWLEY, Andrew

- 1972 On the internal relationships of Eastern Oceanic languages. In: R.C. Green and M. Kelly, eds. *Studies in Oceanic culture history* vol.3:1-142. Honolulu: Bishop Museum.

- 1977 The verb phrase in Proto-Oceanic. MS.

PAWLEY, Andrew and Lawrence A. REID

- 1976 The evolution of transitive constructions in Austronesian. *WPLUH* 8/2:51-74.

REID, Lawrence A.

- 1978 Problems in the reconstruction of Proto-Philippine construction markers. In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*. Fascicle 1: *Western Austronesian*, 33-66. PL, C-61.

- 1979 Evidence for Proto-Philippine nominative marking. *Philippine Journal of Linguistics* 10:1-20.

SCHACHTER, Paul and Fe T. OTANES

- 1972 *Tagalog reference grammar*. Berkeley: University of California Press.

SKEAT, W.W.

- 1900 *Malay magic*. New York: Dover Publications.

STOKHOF, W.A.L., ed., ... with Lia SALEH-BRONCKHORST and Alma E. ALMANAR

- 1981 *Holle lists: vocabularies in languages of Indonesia*, vol.3/2: *Central Moluccas: Seram (II)*. [MLIn 10]. PL, D-44.

STRESEMANN, Erwin

- 1918 *Die Paulohisprache: ein Beitrag zur Kenntnis der Amboinischen Sprachengruppe*. The Hague: Martinus Nijhoff.

- 1927 *Die Lauterscheinungen in den ambonischen Sprachen*. *Zeitschrift für Eingeborenen-Sprachen*, Beiheft 10.

TUUK, H.N. van der

1971 *A grammar of Toba Batak*. Translated by Jeune Scott-Kemball from his *Tobasche spraakkunst*, 2 vols, 1864 and 1967. The Hague: Nijhoff.

VEEN, H. van der and J. TAMMU

1972 *Kamus Toraja-Indonesia*. Rantepao: Jajasan Perguruan Kristen Toradja.

WARNECK, Johannes

1977 *Toba-Batak-deutsches Wörterbuch*. The Hague: Nijhoff.

WILKINSON, R.J.

1959 *A Malay-English dictionary*. London: Macmillan.

ZAINUDDIN UNTUNG

1979 *Keluargaan bahasa-Bahasa di Sulawesi Selatan: Bahasa Indonesia dan Bahasa Bonerate dalam perbandingan seratus kata*. MS.

ZORC, R. David

1982 *Where, O where, have the laryngeals gone? Austronesian laryngeals re-examined*. See pp.111-144 in this volume.

ZURBUCHEN, M.S.

1976 *Introduction to Old Javanese language and literature: a Kawi prose anthology*. The Michigan series in South and Southeast Asian languages and linguistics, 3. Ann Arbor.

THE DEMISE OF PROTO-PHILIPPINES

Lawrence A. Reid

0. INTRODUCTION

This paper constitutes a report of research in progress. Its results are tentative, in that I am still looking for confirmatory evidence, but the general direction of the research seems justified given the evidence presently available. For a number of years I have operated on the assumption that the Philippine languages (and possibly others, external to the Philippine archipelago) constitute a subfamily within the Malayo-Polynesian, or extra-Formosan languages. Certainly the languages are typologically similar, to the extent that the use of the term 'Philippine-type' is widely accepted as defining a certain kind of verbal syntax. It has been a useful assumption too, in that it has enabled us to reconstruct some features of an early proto-language, to which the label Proto-Philippines has been applied.

Clinging to the concept of the Philippines as a subfamily, with each of its languages more closely related to one another than to any language outside the Philippines, has, for me anyway, been a reaction to the completely untenable theory that the different ethnic groups in the Philippines are the result of a series of migrations from the south and the west, a view that was popularised by the late H. Otley Beyer (1948) and which is still taught as fact in Philippine schools today. It is a theory, however, which has never been substantiated by any evidence, linguistic or archaeological. (For a critique of the wave migration theory of Philippine ethnic groups from the archaeologists' and anthropologists' point of view, see Evangelista (1967), Jocano (1967), and Yengoyan (1967).)

If the theory of multiple migrations were true, it would imply that the closest linguistic relatives of each group in the Philippines would be with the group from which it separated at the time of its migration. We ought to be able to trace shared innovations which would establish the unity of each migrant group with the group at its point of origin, in much the same way as the ancestors of the Malagasy-speaking people in Madagascar have been shown to have migrated from Borneo more than fifteen hundred years ago and to be most closely related to the Maanyan language (Dahl 1951). But such evidence for the Philippine languages has not been forthcoming.

The opposite point of view is that all Philippine languages developed in situ, and are daughters of a single parent language which could be called Proto-Philippines. Some attempt has been made to reconstruct a Proto-Philippines, but

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 201-216. *Pacific Linguistics*, C-75, 1982.
© Lawrence A. Reid

there are very few features of this proto-language which can be shown to be innovations restricted only to languages within the Philippine archipelago or its environs. With a situation such as this, can we continue to cling to the notion of a Philippine subfamily, or should we revise our assumptions? Reid (1978), after noting the weak nature of the evidence for a Proto-Philippines, commented, "It is possible that in the future we may have to reassess completely the evidence for a Proto-Philippines, even comprising the languages of the geographical Philippines."

This paper is the beginning of the reassessment. It will give evidence which suggests that the northern languages of the Philippines do not form a sister subgroup with the rest of the Philippines, but with a much larger subgroup, comprising most if not all of the rest of the Malayo-Polynesian family. It will claim that Dempwolff's 'facultative' nasal, which forms nasal clusters in many word bases in Central and Southern Philippine languages, Indonesian and Oceanic languages is an innovation which postdated the separation of the parent of the Northern Philippine languages from the parent of the rest of the extra-Formosan languages. It will also claim that forms with medial nasal clusters which appear in the northern Philippine languages, and which appear to be cognate with forms with a nasal in languages outside this group are borrowed forms.

1. IS THERE ANY EVIDENCE FOR A PROTO-PHILIPPINES?

1.1. There has been little systematic attempt to reconstruct a Proto-Philippines (PPh), apart from the work of Charles (1974), which deals with phonological reconstruction, two papers which discuss the reconstruction of grammatical markers (Reid 1978, 1979), and an unpublished list of possible Proto-Philippine lexical items (Zorc 1971).

Charles did not restrict himself to the comparison of languages of the geographical Philippines, but included, as well, languages of the northern Celebes (Sangirese, Tontemboan, Mongondow and Gorontalo) and northern Borneo (Dusun and Murut), since from the time of Brandstetter (1906), these languages have been recognised as bearing phonological, lexical and syntactic similarity to the languages in the Philippines.

The phonological developments which Charles proposed for Proto-Philippines from Proto-Austronesian were as follows:

- a. PAN *C, *T and *t do not contrast in PPh, they are reflected simply as PPh *t.
- b. PAN *c and *s merged as *s in PPh.
- c. PAN *d, *D, *z and *Z similarly have merged to PPh *d.
- d. PAN *-ey and *-ay merged in PPh *-ay.
- e. None of Dyen's subscripted PAN phonemes (R₁-R₄, S₁-S₅, etc.) or the other PAN phonemes reconstructed by Dyen to account for variant Formosan reflexes are distinguished in the Philippines.

Two problems arise. One, is that some of Dyen's and Dempwolff's proposed phonemic contrasts for PAN do not stand close scrutiny, and are probably the result of unrecognised borrowing or obscured phonological processes in the history of the languages involved. (Wolff, in his paper in this volume, challenges the reconstruction of PAN *c, *z, *g, and *T.) The other problem is that the suggested mergers are by no means unique to the Philippines. All languages outside of Formosa have merged *C and *t. All but a few western Indonesian

languages (Javanese, Malay, Acehnese and Malagasy) have also merged the reflexes of *c and *s. Only Javanese and Madurese distinguish retroflexed apicals (*T and *D) from non-retroflexed (*t and *d). The status of *z as distinct from *Z is questionable, and there are many languages outside the Philippine group that do not have a distinct reflex of *Z from *d and *D. None of the mergers which were listed by Charles are innovations which exclusively appear in the Philippines, either in its restricted geographical sense, or in its extended typological sense.

Charles (1974:487) stated that the reason he included the languages of north Borneo in the Philippine subgroup is that they merge *z, *Z, *d, *D and *j, implying that this is a merger which characterises the Philippines. The Cordilleran languages of the Philippines however (as Charles also notes) do not merge *j with *d. Their reflex of *j is generally g. Blust (1974) furthermore argues that the northern Borneo languages do not subgroup with languages to the north of it (in the Philippines) but to languages further south in Sarawak, on the basis of an exclusively shared innovation in those languages by which a vowel was deleted when it occurred between reflexes of a PAN voiced obstruent and a following *S, with subsequent change of *s to h.

1.2. The articles which attempted to reconstruct grammatical forms for Proto-Philippines (Reid 1978, 1979), can be examined under the same light as Charles' paper on Proto-Philippine phonology. Although the forms which were reconstructed provide a plausible proto-system, and account for much of the diversity found in grammatical markers in Philippine languages, nothing is proposed which could not be present in an earlier proto-language, one that would perhaps be the parent of all extra-Formosan languages. There are no exclusively shared innovations separating the Philippine languages either in the restricted or the expanded sense, from other non-Philippine languages.

1.3. Zorc (1971) is a compilation of reconstructed forms from the works of Dempwolff, Dyen and Blust, which have reflexes in some Philippine languages. These are supplemented by reconstructions by Zorc and by Charles based on a comparison of Philippine lexicons. If there were a Proto-Philippines, the items in this list would probably have formed a fair part of its vocabulary. However, they do not themselves give evidence for a Proto-Philippines, as evidenced by Zorc's tongue-in-cheek title (Proto-[Proto-(Proto)]-Philippine finder list).

1.4. The Philippines has also been considered a single subgroup because of the 'focus' system of syntax. However a similar system is also found in Formosa (especially in Amis) as well as in languages to the south of the Philippines, and in Malagasy. Chamorro also shows evidence that it, too, once had such a system. All of these systems evolved from a Proto-Austronesian syntax, described in detail in the paper by Starosta, Pawley and Reid in this volume. Whether the verbal focus systems in non-Philippine languages developed independently from the systems that developed in the Philippines, or whether a developed verbal focus system evolved only once in the history of Austronesian is still an unanswered question. We still cannot say with assurance that the

possession of a verbal focus system is a shared innovation which unites those languages that have one into a single subfamily of Austronesian.

There is no evidence which we can at present produce to support the hypothesis that Philippine languages form a single subgroup within Austronesian and until there is, we probably should stop talking about a Proto-Philippines.

1.5. A number of works have noted the aberrant nature of certain of the Philippine languages, and have suggested that these languages do not constitute a part of a Philippine subgroup. It has usually been the low lexicostatistical percentages that these languages have when compared to other Philippine languages that has justified their being excluded from the Philippine subgroup. These languages include Ivatan, Ilongot, Tiruray, Bagobo, and the Bilic subgroup (Blaan and Tboli). At the present time, Ilongot is the only one of these languages which can with any assurance be subgrouped within an accepted Philippine subgroup — Southern Cordilleran (Reid 1979), also called Pangasinic (Zorc 1979). But even this assignment of Ilongot is challenged by Walton (1979) and McFarland (1980).

The Bilic languages are a different story. They are not just lexically divergent. They are phonologically, morphologically and syntactically divergent. Their position in relation to other Philippine languages is particularly suspect. I will return to these languages again below.

The genetic affiliation of the Sama-Bajaw languages is also questionable. Pallesen (1977:339) states,

The geographical origin and immediate linguistic affiliates of PSB [Proto-Sama-Bajaw] have not yet been determined. A number of distinctive characteristics (e.g. the h reflex of PAN *R, the semantic features of the phrase marking particles (or prepositions), the lack of verbal inflection to mark the action-begun vs. action-not-begun contrast, a 7-vowel system, a uniquely marked agentive phrase) indicate an Indonesian origin rather than any close relationship to the Central Philippine languages with which many SB daughter languages are currently in geographical proximity.

2. MEDIAL NASAL CLUSTERS, AND THE NORTHERN PHILIPPINE LANGUAGES

The position of the northern languages of the Philippines, in particular the Cordilleran group vis-a-vis the languages in the central and southern areas of the Philippines has never to my knowledge been seriously questioned. It has been assumed by everyone doing Philippine comparative work that these represent the two major divisions of Philippine languages, and cognates appearing in both groups would automatically qualify for inclusion in a Proto-Philippines.

2.1. The case of Bontok

I began to question the status of the genetic relationship between the Cordilleran languages and the rest of the Philippine group in 1978, when a student (Mary Nutthal, Department of Anthropology, Auckland University) in a paper in which she was identifying the reflexes of reconstructed lexical items in Bontok, a Central Cordilleran language, reported that although she had identified a large number of reflexes of supposedly PAN and PPh words, there were relatively few apparent reflexes of the many items in Dempwolff (1938), Blust (1970) and Zorc (1971) which had been reconstructed with either an obligatory, or optional prenasalised medial consonant.

The forms which were possible reflexes generally showed either phonological evidence of having been borrowed (as in Table 1), or did not have a medial nasal (as in Table 2). The following reconstructions are cited from Zorc (1971). The Bontok forms are from Reid (1976).

Table 1: Borrowed medial nasal cluster forms in Bontok showing irregular phonological reflexes

*baNkaw 'spear, lance'	>	bangkaw (expected bangkew)
*baNtay 'guard, watch'	>	bantay (expected bantey)
*saNlaR 'toast, fry'	>	sanglag (expected sanglal)
*t+uNduR 'follow'	>	tundug (expected tundul)

Table 2: Bontok reflexes which do not show a medial nasal cluster

*aNpil 'favour one person over another'	>	qapil
*a(N)pu 'grandparent, child'	>	qapu
*a(N)taq 'unripe'	>	q:ata
*i(N)pit 'squeeze'	>	qi:pit
*u(N)tek 'brain'	>	qu:tek
*buNkar 'toss around'	>	bu:kal
*DeNpilas 'cliff'	>	deplas
*ke(N)pes 'shrink, shrivel'	>	kepes
*ke(N)pit 'press, clamp'	>	kipit
*kiNdat 'open up eyes'	>	ki:dat
*la(N)pis 'layer, thin, slab'	>	lapisdak
*leNpit 'fold'	>	ka-lpit
*li(N)kep 'shut a door'	>	likep
*li(N)kung 'concave, bent'	>	likung

<i>Table 2 (cont)</i>	
*lu(N)tuq 'cook'	> lu:tu
*ciNduk 'ladle, scoop'	> si:duk
*su(N)wan 'tool'	> su:wan
*ta(N)bun 'cover up'	> ta:bun
*ta(N)gi 'ask payment'	> ta:gi
*ta(N)pi[] 'winnow'	> t-in-api
*ti(N)pun 'assemble, unite'	> ti:pun
*tu(N)buq 'grow, increase'	> tu:bu

Several questions arose as a result of this discovery that Bontok did not appear to reflect the nasal in forms with reconstructed medial nasal clusters.

2.1.1. One question was, could Bontok have lost the nasal in such forms? The evidence would seem to indicate that the answer to this question is negative. There is no synchronic or other evidence which would indicate a phonological restriction on medial nasal clusters. To the contrary, all possible nasal plus consonant clusters occur in the language, both homorganic and heterorganic, as the result of loss of an unstressed pepet vowel following either -in- or -um- infixes, and the introduction of borrowed forms with such clusters. Neither is there any evidence to suggest that medial nasal clusters developed as non-nasal clusters as has been proposed by Blust (1980) for the Chamic languages, where such clusters may have become sequences of glottal stop plus consonant.

2.1.2. Another question which needed to be answered was, to what extent is the situation in Bontok typical of the northern languages of the Philippines? This question was not as easy to answer as it was for Bontok, partly because of the dearth of good dictionaries, and partly because sound changes in the Northern Cordilleran languages tend to obscure borrowed forms. In all of those languages (except Ilokano), the regular reflex of *R is g, as it is in Tagalog and the other Central Philippine languages. Similarly, *a before *y or *w does not become e, as it does in the Central and Southern Cordilleran languages, but remains a as it does in Tagalog. It was apparent however that there were far fewer possible reflexes of forms with a medial nasal cluster than there were of forms without the cluster. Furthermore, those that did occur were often the same forms as those which could be shown to be borrowings in Bontok.

The northern Philippine language with the greatest number of possibly inherited forms with a medial nasal cluster is Ilokano. However we can show that a large proportion of these forms are probably borrowings. To this we will return below.

2.1.3. A third question arose as the result of the realisation that Bontok probably did not have reflexes of reconstructed medial nasal clusters. The origin of the medial nasal in the forms in question has been considered by Dahl (1973), Latta (1978) and others to be the same as the final nasal in the *maN-* prefix. If this be so, how could one explain the fact that Bontok has a reflex of **maN-*, but no medial **-N-*? The answer to this is that whether or not there was a stage in the history of those languages when the infixation of **-N-* into root words functioned in the same way as infixation of **-N-* into **ma-* derived forms, viz. as a transitivising affix, the phonological processes associated with each are different. This has long been recognised and has been extensively discussed in the literature. Both the medial nasal and the *maN-* final nasal assimilate to a following non-continuant. It is only *maN-*, or *N-* at the beginning of a word which triggers nasal substitution (or deletion) of the following consonant. Infixation of *-N-* into a root, produced assimilation, but no deletion. If the development of these nasal infixes occurred at the same time in the history of the languages, it is difficult to account for the difference in phonological processes associated with each.

One other set of facts is of crucial importance here in determining the relative chronology of the development of these two nasal infixes. In Javanese, *N-* (from **maN-*) assimilates to *n'* before *s* < **t'*. The assimilation to a palatal nasal reflects the fact that the following *s* was originally **t'*, a palatal non-continuant. The medial nasal **-N-*, however, does not assimilate to a following *s* < **t'*, but remains *ng*. Dempwolff claimed that Javanese *-ngs-* was the result of dissimilation, but both Blust (1970) and Dahl (1973) have argued that the Javanese medial *-ngs-* sequence can be better accounted for by claiming that the assimilation of **-N-* to a following non-continuant only began to operate after **t' > s*, so that Javanese *-ngs-* never fit the structural description necessary for assimilation to take place. Latta (1978) states that Javanese *n'* for initial *s* creates a relative chronology problem for this hypothesis, one for which he could see no obvious solution. It would seem that the only reasonable explanation for these facts is that infixation of **-N-* into root words was subsequent to the change **t' > s*. The assimilation was not triggered because *s* is not a non-continuant.

There is no problem then, in having a language such as Bontok in which reflexes of **maN-* occur but no reflexes of **-N-*. Infixation of **-N-* into root words must have postdated the separation of the parent of Bontok from the parent of the languages in which reflexes of **-N-* appear.

2.2. The case of Ilokano

There are around 50 words in Ilokano which contain a medial nasal cluster which could possibly be a reflex of a reconstructed nasal cluster. If the thesis of this paper is correct, none of these forms is inherited, because the process by which medial nasals became infixed developed after the separation of the northern languages of the Philippines from the rest of the extra-Formosan languages.

That Ilokano has been heavily influenced by outside contacts, should come as no surprise. Ilokanos, being a coastal people, were as much involved in trade with whomever passed their shores as people further south were. In addition to coastal trade, Ilokanos probably also developed overland trade routes through Pangasinan and Kapampangan territory to Manila.

Pallesen (1977) has documented the routes of Sama-Bajaw traders from the Sulu area into the Bisayas close to a thousand years ago, and has also documented some of their probable settlements in that area. It would seem strange if having gone that far north and establishing trading bases, they did not also sail up the west coast of Luzon to tap that region for the wealth of forest products and possibly gold that was available in the Ilokos region. Such trips would almost certainly have required refitting and repair stops, or stopover periods during the bad weather months in the second half of the year when sailing would have been dangerous.

Pallesen (1977:362) discusses the linguistic evidence for the establishing of such a settlement point at the Agusan River mouth in north-east Mindanao. He raises, for example, an unanswered question regarding the source of the Northern Sama (q)ag- actor focus prefix. It may be coincidental that Ilokano seems to be the only other Philippine language that uses qag- with this function. It may also be coincidence that Sama-Bajaw and Ilokano share the phonological process of consonant gemination after a pepet vowel. But these facts may gain significance if further linguistic evidence of contact can be demonstrated.

Pallesen suggests that the presence of the form daqing meaning '*jerked fish*' in Ilokano, Tagalog and Malay is evidence of Sama-Bajaw contact, since daqing is the generic term for fish in this language, and its export from Sulu as '*jerked fish*' would account for its being borrowed with this meaning into the languages of the peoples with whom the Sama-Bajaw traded.

Wolff (1976) has also presented a convincing case for a sizable, and influential, Malay-speaking community in the Manila area prior to European contact, and has documented scores of loans from Malay which came into Tagalog as a result. Many of these forms (both with and without medial nasal clusters) are also found in Ilokano and neighbouring languages, having been borrowed directly from Malay, or indirectly through Tagalog.

The primary evidence that these are loanwords is phonological. Wolff has charted a number of Javanese-Malay-Tagalog correspondences (Wolff 1976:351) which are indicative of borrowed terms in Tagalog. Where there is an Ilokano word which is identical to one in Tagalog which has been shown to be a borrowing, it is assumed that Ilokano has borrowed the Malay word via Tagalog. Sometimes an Ilokano word phonologically corresponds more closely to the Malay form than to the Tagalog, and in such cases direct borrowing from Malay is assumed. An example is the now obsolescent Ilokano word for coat or shirt, badio (Tagalog bado), which was without doubt a direct borrowing of Malay bad'u.

Considering the likelihood that Ilokano trade was probably carried on with itinerant Sama-Bajaw seafarers, we cannot rule out the possibility that some Ilokano forms (such as tamban '*sardine*', which is identical in Tagalog, Malay and Sama-Bajaw) may have been borrowed directly from Sama-Bajaw rather than from Malay.

There are a number of Ilokano forms with medial nasal clusters for which no direct phonological evidence can be cited to prove borrowing. Nevertheless their meanings are suggestive of borrowed forms. Words associated with trade are prime candidates for borrowing. Such items would have included not only terms for objects that were commonly traded, but also the ways in which those objects were measured or counted, and the ways in which trading was commonly conducted. Other semantic clues for borrowing are found in words which refer to cultural adaptations, such as certain kinds of clothing (which I assume

were introduced after the first Austronesian settlement in Northern Luzon), and words for church and street (both probably relatively recent borrowings). The names for fruits and plants are also very susceptible to borrowing, and we find a number of these in the list of suspected borrowings.

The list of Ilokano words in Table 3 includes those forms having a medial nasal cluster which are suspect of being borrowed because of a phonological irregularity, or which have a non-Austronesian source.

Table 3: Borrowed medial cluster forms in Ilokano which show phonological irregularity

qampir ' <i>sign, token</i> '	< Ml ampir ' <i>near</i> '
bibingka ' <i>rice cake</i> '	< Ml bingka, via Tg bibingka
kambal ' <i>double yolk</i> '	< Ml kembar, via Tg kambal
lampin ' <i>diaper</i> '	< Ml romping-ramping ' <i>rag</i> ', via Tg lampin
lansa ' <i>nail, peg</i> '	< Ml ran't'ang, possibly via Tg. Note Ceb lansang
langkuas ' <i>Alpinia galanga</i> '	< Ml lengkuas, via Tg langkwaw
lumba ' <i>race</i> '	< Ml lumba, or Tg lumba
lumbalumba ' <i>marine fish</i> '	< Ml lumbalumba ' <i>dolphin</i> ', also Tg
angka ' <i>jackfruit</i> '	< Ml angka, or Tg angkaq
palangka ' <i>movable seat</i> '	< Ml palangka, also Tg ' <i>sedan chair</i> '
panday ' <i>blacksmithing</i> '	< Ml panday ' <i>skillful</i> ', via Tg panday ' <i>blacksmith</i> '
pinggan ' <i>dish</i> '	< Ml pinggan ' <i>bowl</i> ', via Tg pinggan ' <i>dish</i> '
rangkap ' <i>gift, tip, alms</i> '	< Ml rangkap ' <i>pair</i> ', Tg langkap ' <i>that which is joined to something</i> '
ranggas ' <i>damage, spoil</i> '	< Ml ragas, ranggas ' <i>defoliated</i> '
sampaga ' <i>Arabic jasmine</i> '	< Ml t'ampaka, via Tg sampaga
sampan ' <i>boat</i> '	< Ml, Tg sampan < Chinese
sampor ' <i>reel of twine of two types</i> '	< Ml t'ampor ' <i>mixed</i> '
santol ' <i>k.o. tree</i> '	< Ml sentol, via Tg santol
simbaan ' <i>church</i> '	< Ml sembah ' <i>respect</i> ', via Tg simbahan ' <i>church</i> '
simpan ' <i>put in order</i> '	< Ml simpan ' <i>be ready, have finished</i> ', or Tg simpan ' <i>equipment</i> '
tanda ' <i>sign, token</i> '	< Ml tanda, or Tg tandaq
tanggigi ' <i>Spanish mackerel</i> '	< Tg tanggingiq, tangigiq, or tanigiq
timba ' <i>pail</i> '	< Ml timba, or Tg timbaq
timbang ' <i>scales</i> '	< Ml, Tg timbang
tunda ' <i>lead</i> '	< Ml tunda ' <i>be towed</i> ', or Tg tundaq

The following table of Ilokano words includes those which although their status as borrowed words cannot be verified by phonological means, are nevertheless suspect of being borrowed because of their semantics. Their possible sources in Malay and Tagalog are cited.

Table 4: Ilokano words with medial nasal clusters suspect of being borrowings from Malay or Tagalog

qingga 'end'	Ml hingga
bangkudo 'red cotton'	Ml bengkudu, Tg bangkuro 'k.o. tree used for red dye'
bunton 'heap, pile'	Tg bunton
dampag 'breadth'	source uncertain
gimbal 'drum'	Tg gimbal
lansangan 'street'	Tg lansangan
mangga 'mango'	Tg mangga
pandek 'short'	Tg pandak
pandong 'mantilla, veil'	Tg pandong
salambaw 'k.o. fishing net'	Tg salambaw
sambot 'redeem, repay, recover'	Tg sambot 'catch in the hand'
sangga 'parry, ward off'	Tg sangga
singked 'ratify, confirm'	Tg singkad 'full, exact, complete'
sumpit 'syringe, blowpipe'	Tg sumpit 'blowgun, enema'
tamban 'sardine'	Tg tamban
tandok 'surgical cupping instrument sometimes made of horn'	Tg tandok, cf. Ml tandok 'horn'. Wolff considers the semantic similarity between the Tg and Ml forms as fortuitous.
tangkay 'stem, stalk, umbrella, handle, tool handle'	Tg tangkay
tumbok 'iron pestle, melt and mould metal'	Tg tumbok 'strike with pointed object'
tunggal 'each, every'	Tg tunggal 'buy and sell one by one'

The only remaining forms in Ilokano which contain a medial nasal cluster and which may be reflexes of a reconstructed form with such a cluster are given in Table 5.

Table 5: Additional Ilokano words
with medial nasal clusters

qandap	<i>'luminescence'</i>
balantik	<i>'to flick'</i>
bantay	<i>'guard'</i>
bangkay	<i>'corpse'</i>
(g)ampang	<i>'frivolous'</i>
lindong	<i>'shade'</i>
sanggir	<i>'lean, incline'</i>
tambak	<i>'dam'</i>

Of these forms, *bantay* and *bangkay* occur identically in some of the Central and Southern Cordilleran languages in which, were they inherited, they would have final *-ey* for **-ay*. It would be a reasonable inference that if they were borrowed in the Southern and Central Cordilleran languages, they were also borrowed in the Northern Cordilleran languages but without a phonological trace. These few remaining forms do not belong to the area of basic vocabulary, and consequently may well be borrowed. Certainly we would not wish to establish the occurrence of medial nasal clusters in the parent of Ilokano on the basis of such forms.

3. CONCLUSION

If, as the evidence suggests, the northern languages of the Philippines do not share in nasal infixation into root words, the implications for subgrouping are fairly clear. Those languages which do share it form a subgroup within Austronesian, and the northern languages of the Philippines are not part of that subgroup.

The Formosan languages are generally considered not to have reflexes of an assimilating nasal infix, either into root words or into *ma-* prefixed words. Benedict's (1976) attempt to identify such forms in Formosan languages is quite unconvincing. The forms which do show medial nasal clusters in Formosan languages can be accounted for by assigning the nasal to a reflex of PAN **-um-* or **-in-*.

A careful evaluation of the status of medial nasals in other Philippine languages is also called for. *Blaan* and *Tboli*, for example, do not appear to share in medial nasal infixation, nor is there any evidence that they ever had a *maN-* prefix. In this respect they are unlike most other extra-Formosan languages. They are unique also in that these are the only Philippine languages which use a reflex of **-in-* solely to mark 'object focus' and not also as an aspect marker. Their antiquity is revealed also in their phonology. *Tboli* is the only Philippine language (apart from *Tagbanwa*) which retains PAN **q* as a backed velar stop. Other Philippine languages have glottal stop (and zero) as their reflex of **q*.

It is possible that Blaan and Tboli provide evidence of a very early migration south of Formosa by an Austronesian-speaking people. The northern languages of the Philippines are assumed to have developed from a separate migration, one which originated from the Amis-speaking area in south-east Formosa and which possibly passed by way of Botel Tobago, leaving a Yami-speaking population, to Y'Ami island and the Batanes archipelago, before moving south to populate the northern areas of Luzon. It is significant that the name Amis appears to be cognate with Tagalog qamihan, Ilk qamian '*north wind*', Ilk qamianan '*north*'. Such formal similarities lend themselves to speculation as to the direction of the semantic development. If the Amis-speaking area was the dispersal point for Austronesian speakers to the south it is possible that *amis-an '*place of the Amis*' could have become associated with the meaning '*north*'. On the other hand if migration had proceeded in the opposite direction, from south to north, it seems unlikely that this particular semantic development could have taken place.

A movement south from the northern Philippine area, resulted in the development of a subgroup within which medial nasal infixation developed. It is convenient to refer to this latter group as Malayo-Polynesian, and to include within it not only all the Austronesian languages of Indonesia and Oceania, but at least the Central Philippine languages as well.

The position of some of the other languages of the Philippines such as the Mangyan groups, the Manobo and Danao groups, Tiruray and Bagobo is uncertain. If it can be shown that forms with medial nasal clusters in these languages are the result of borrowing, and are not directly inherited, then they will need to be removed from the Malayo-Polynesian subgroup. As one moves south in the Philippines however, 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.

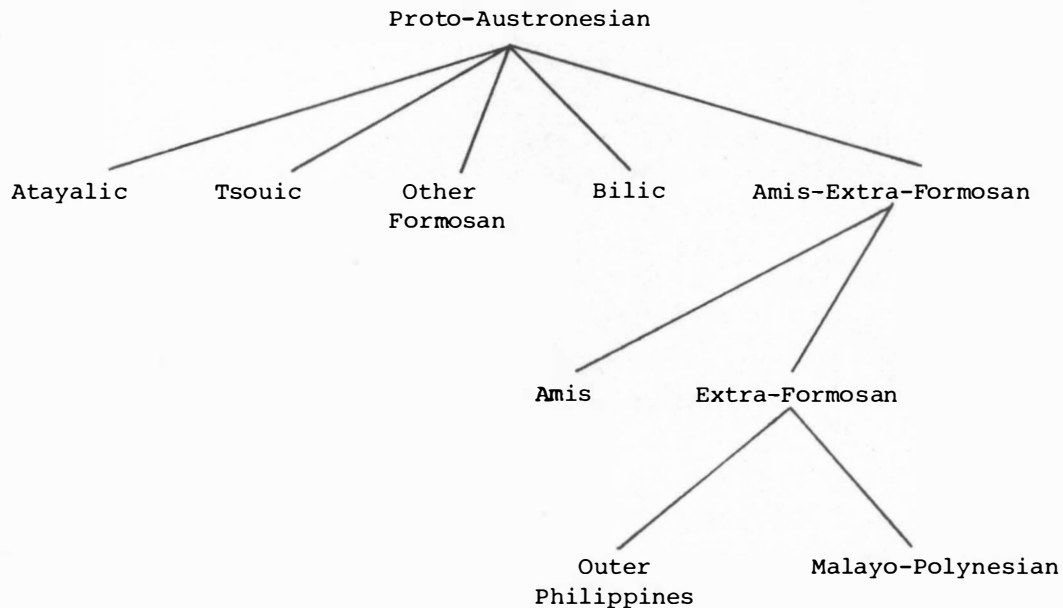
We must remember also that we are talking about contact between possibly geographically adjacent languages, which must have been going on for seven thousand years or more. The dispersal of the Oceanic group is gauged by Pawley and Green (1973:52) to have begun not later than 5000 years ago. They cite archaeological evidence that New Caledonia has been occupied since 3000 B.C. If Formosa was indeed the homeland of Proto-Austronesian the dispersal of the Austronesians could hardly have started later than 5000 B.C.

An attempt has been made in this paper to show that languages which share medial nasal clusters as the result of an infixed *-N-, form a genetic subgroup distinct from those which do not. Since Philippine languages are split between those which do and those which do not, and those which do share the nasal infixing innovation with Indonesian and Oceanic languages, we can no longer consider Philippine languages as constituting a single subgroup within the Hesperonesian languages.

Other possible innovations may also be identified for this redefined Malayo-Polynesian, which may make it easier to definitively exclude some if not all southern Philippine languages and possibly some languages to the south of the Philippines from it. One such possible innovation which may be useful for this purpose is the development of a velar nasal variant of the PAN ligature *a following vowels, with the ultimate development of nasal final determiners such as Tagalog ang, Javanese sang, etc., (Reid 1978), an innovation which may not be unrelated to the nasal infixation process which has been discussed in this paper. Because of the likelihood that evidence will be forthcoming to include southern Philippine languages with those of the northern Philippines in

a subgroup distinct from the central Philippine languages, I now use the term Outer Philippines to label all non-Malayo-Polynesian Philippine languages.

The subgrouping of PAN at its highest level, which best accounts for the above facts is a modification of that proposed by Blust, and by Harvey:



BIBLIOGRAPHY

BENEDICT, Paul K.

- 1976 Formosan reflexes of PAN nasal/orals. In: Nguyễn Đăng Liêm, ed. *South-East Asian linguistic studies*, vol.2:237-251. PL, C-42.

BEYER, H. Otley

- 1948 Philippine and East Asian archaeology and its relation to the origin of the Pacific Islands population. *Bulletin of the National Research Council of the Philippine Islands* 29:1-130. Manila.

BLUST, Robert A.

- 1970 Proto-Austronesian addenda. *Oceanic Linguistics* 9/2:104-162.
 1974 The Proto-North Sarawak vowel deletion hypothesis. Ph.D. dissertation, University of Hawaii.
 1980 More on the origins of glottalic consonants. *Lingua* 52:125-156.

BRANDSTETTER, Renward

- 1906 Die Stellung der minahassischen Idiome zu dem...übrigen Sprachen von Celebes einerseits und...zu dem Sprachen der Philippinen anderseits. Quoted in Fritz Sarasin *Materialen zur Naturgeschichte der Insel Celebes*, vol.5: *Versuch einer Anthropologie der Insel Celebes*, part 2: *Die Varietäten des Menschen auf Celebes*, 35-38. Wiesbaden: C.W. Kreidel.

CHARLES, Mathew

- 1974 Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. In *Papers of the First International Conference on Comparative Austronesian Linguistics, Honolulu, 1974*, 2. *Oceanic Linguistics* 13:457-509.

DAHL, Otto Christian

- 1951 *Malgache et Maanjan: une comparaison linguistique*. Oslo: Egede-Instituttet.
 1973 *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series, 15. Lund: Studentlitteratur.

DEMPWOLFF, Otto

- 1938 *Vergleichende Lautlehre des austronesischen Wortschatzes*, vol.3: *Austronesisches Wörterverzeichnis*. Berlin: Dietrich Reimer.

EVANGELISTA, Alfredo E.

- 1967 H.O. Beyer's Philippine Neolithic in the context of postwar discoveries in local archaeology. In: Zamora, ed. 1967:63-87.

JOCANO, F. Landa

- 1967 Beyer's theory on Filipino prehistory and culture: an alternative approach to the problem. In: Zamora, ed. 1967:128-150.

LATTA, Christian

- 1978 Internal reconstruction. Ph.D. dissertation, Ohio State University.

McFARLAND, Curtis D.

- 1980 *A linguistic atlas of the Philippines*. Monograph 10. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

PALLESEN, A. Kemp

- 1977 Culture and language convergence. Ph.D. dissertation, University of California, Berkeley.

PAWLEY, Andrew and Roger C. GREEN

- 1973 Dating the dispersal of the Oceanic languages. In *Papers of the First International Conference on Comparative Austronesian Linguistics, 1974 - Oceanic*. *Oceanic Linguistics* 12:1-67.

REID, Lawrence A.

- 1976 *Bontok-English dictionary* [with English-Bontok finder list]. PL, C-36.
- 1978 Problems in the reconstruction of Proto-Philippine construction markers. In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 33-66. PL, C-61.
- 1979 Evidence for Proto-Philippine nominative marking. *Philippine Journal of Linguistics* 10:1-20.

STAROSTA, Stanley, Andrew K. PAWLEY and Lawrence A. REID

- 1982 The evolution of focus in Austronesian. See pp.145-170 in this volume.

WALTON, Charles

- 1979 A Philippine language tree. *Anthropological Linguistics* 21/2:70-98.

WOLFF, John U.

- 1976 Malay borrowings in Tagalog. In: C.D. Cowan and O.W. Walters, eds *Southeast Asian history and historiography: essays presented to D.G.E. Hall*, 345-367. Ithaca, N.Y.: Cornell University Press.
- 1982 Proto-Austronesian *c, *z, *g and *T. See pp.1-30 in this volume.

YENGOYAN, Aram A.

- 1967 The initial populating of the Philippines: some problems and interpretations. In: Zamora, ed. 1967:175-185.

ZAMORA, Mario D., ed.

- 1967 *Studies in Philippine anthropology (in honor of H. Otley Beyer)*. Quezon City: Alemar-Phoenix.

ZORC, R. David

- 1971 Proto-[Proto-(Proto)]- Philippine finder list. Xerox, Cornell University.
- 1979 On the development of contrastive word accent: Pangasinan, a case in point. In: Nguyễn Đăng Liêm, ed. *South-East Asian linguistic studies*, vol.3:241-258. PL, C-45.

THE HALMAHERA CONNECTION:
A CASE FOR PREHISTORIC TRAFFIC THROUGH TORRES STRAITS

C.L. Voorhoeve

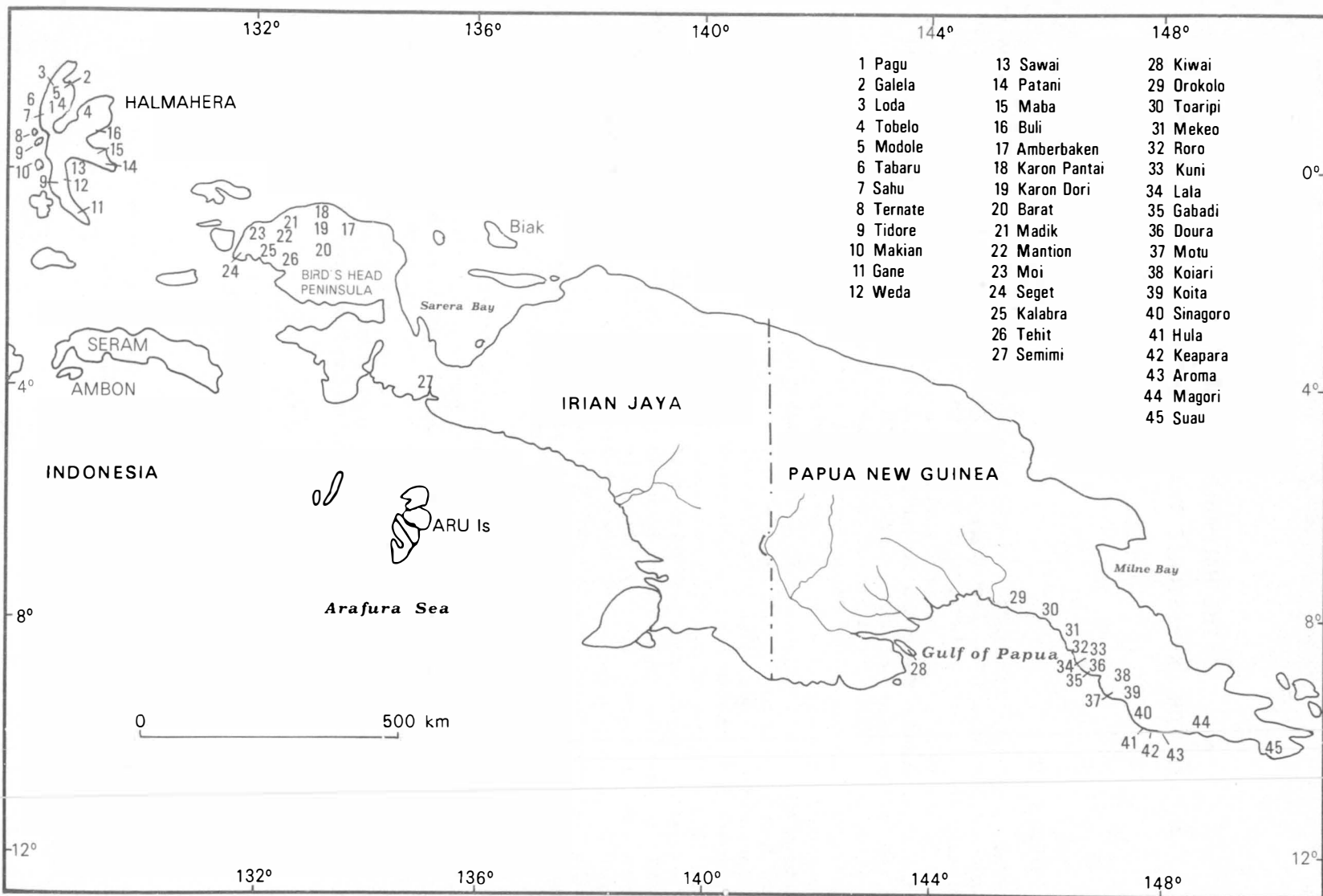
1. INTRODUCTION

This paper¹ is an offshoot from a search for Austronesian (AN) loanwords in the non-Austronesian (NAN) languages of the North Moluccas, motivated by the need to clarify the genetic relationship between those languages and the Papuan languages in the western and central parts of the Bird's Head, Irian Jaya. The relationship between these two distantly related groups of languages has been obscured by the intake of large numbers of AN loanwords into each of them. In the course of the investigation it became clear that apart from loanwords of Indonesian, Moluccan Malay, Ambon-Seram, and Philippine origin, and a number of borrowings from unspecifiable West Austronesian and Oceanic sources there is in the North Moluccan languages – or North Halmaheran (NH) languages as they have become known – a number of AN loans which, as a body, point to the Central Papuan (CP) languages in south-east Papua New Guinea as their possible source. A closer investigation of the CP languages then revealed that they share with the NH languages also a number of probably non-Austronesian cognates. The closeness of the lexical correspondences and their relatively large number (19) rule out the possibility that the CP languages received those words from a Papuan source other than the NH languages. Firstly, the NH languages and the Papuan languages in eastern Papua New Guinea belong to two different phyla, the West Papuan Phylum and the Trans-New Guinea Phylum respectively.² Secondly, that another language group within the West Papuan Phylum would be the source is ruled out by the distant genetic relationship between the NH languages and the West-Central Bird's Head languages.³

In an attempt to account for these facts the hypothesis is here put forward that at a certain time in the far past there has been contact between speakers of a language or languages ancestral to those now spoken in the North Moluccas and ancestors of the speakers of the present-day Central Papuan languages. It will further be argued that this contact took place in south-east New Guinea and must have involved the movement by sea of people from the area west of New Guinea to the east and back via a route roughly following New Guinea's south coast.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics, Tracking the travellers*, 217-239. *Pacific Linguistics*, C-75, 1982.

© C.L. Voorhoeve



The organisation of the paper is as follows: sections 2 and 3 contain short surveys of the North Halmaheran and Central Papuan languages, providing the necessary background information. Section 4 gives an overall discussion of the evidence and some notes on its presentation, which follows in section 5. Section 6 deals with the question of how this contact came about and presents some evidence for the assumption that that it involved traffic along the south coast of New Guinea. The remainder of the paper is taken up by notes on the sources consulted (7), notes (8), an appendix containing phonological and comparative data on the NH and CP languages (9), bibliographical references and a map.

2. THE NORTH HALMAHERAN LANGUAGES

There are ten non-Austronesian languages in the north Moluccan area: Pagu, Modole, Tobelo, Galela, Loda, Tobaru, Sahu, Ternate, Tidore, and West Makian. With the exception of West Makian they are closely related; together they form a stock-level family⁴ within the West Papuan Phylum. The other member languages of the phylum are all found in the western, north-western and central parts of the Bird's Head of Irian Jaya.⁵ Their true relationship to the NH languages has not yet been properly assessed but so much is clear that it is a distant one.⁶ According to Wurm (1975) the languages of the West Papuan Phylum represent a very old Papuan stratum which was already in the area before several of the main Papuan migrations within New Guinea took place.

The nearest Austronesian neighbours of the NH languages are seven small languages spoken on the south coast of Halmahera and on Makian Island: Buli, Maba, Patani, Sawai, Weda, Gimán (Gane), and East Makian.⁷ They belong to the South Halmahera-West New Guinea group, one of the first-order subgroups of Blust's Eastern Malayo-Polynesian.⁸ To the south of Halmahera we find the languages of the Ambon-Seram group, to the north the Philippine languages of which the languages of the Sangir-Talaud subgroup are geographically closest to Halmahera, and in the west the West Austronesian languages of Sulawesi.

3. THE CENTRAL PAPUAN LANGUAGES

The Austronesian languages of the Central Papuan subgroup of the Oceanic languages form a single family located in the Central Province of Papua New Guinea. They fall into at least two subfamilies: Motuic and East-Central Papuan. The Motuic subfamily consists of one language isolate, Motu, spoken along the coast in and immediately east and west of Port Moresby, and the West-Central Papuan languages. There are six of these, located along the coast and further inland immediately west of the Motu-speaking area. They are Mekeo, Roro, Kuni, Lala, Gabadi, and Doura. The East-Central Papuan subfamily has only two members, Sinagoro and the Hula-Keapara-Aroma dialect chain. They are found along the coast about one hundred kilometres south-east of Port Moresby. A third group, whose position within the Central Papuan Family has not yet been cleared up, is found another hundred kilometres further to the east. It consists of four small languages, Magori, Yoba, Bina, and Ouma. Yoba, Bina, and Ouma have only recently been discovered by T.E. Dutton (Dutton 1976).⁹ They show strong lexical influence from the surrounding non-Austronesian languages (see also Dutton 1982).

The Central Papuan Family is one of the thirteen language families in the 'tail' of Papua and the Milne Bay area descended from one proto-language which

has recently been labelled Proto-Papuan Tip (Ross 1979b). Its nearest relative is the Suau Family (Suau, Tubetube) on the south side of the tip of Papua. It is generally accepted that the Austronesians who settled in Central Papua came from the east and that the West-Central Papuan languages represent the westernmost expansion of this migration.

4. THE EVIDENCE

The qualitative evidence which forms the basis of the contact hypothesis consists of 31 sets of probable cognates. They fall into three groups. Those of the first group (section 5, nos 1-12) are of Austronesian origin and occur as loanwords in the NH languages. They all display a feature or features which point to Proto-Central Papuan (PCP) or one of its daughter languages as a possible source. Those of the second group (nos 13-26) probably are non-Austronesian words which have entered the CP languages as loanwords. The third group (nos 27-31) is a residual group: it contains two sets (27,28) of which the CP members could be lexical innovations of PCP, borrowed by the NH languages, and three sets (29-31) which give no clues as to the direction of borrowing.

The main features which link the NH and the CP languages in the sets of group I are:

1. The presence of what seem to be shared innovations: initial g- in set no.1, lo- in set no.3, and loss of POC¹⁰ *r in set no.4;
2. The presence, in the NH forms, of unexpected reflexes of POC consonants which lose their irregularity if one assumes that the source language had undergone sound changes identical or similar to those which occurred in the CP languages. The most obvious of these is NH *h¹¹ < POC *p where one would expect the reflex to be p¹², which gives grounds to assume that the source language had shifted POC *p to some kind of labial fricative, as had happened in PCP (*p) (sets nos 3,4,6,7,11). Others are:

NH *r instead of *t from POC *t (PCP *r).	Set no.2.
NH *r instead of *s from POC *s (PCP *d).	Set no.5.
NH *D instead of *s from POC *s (PCP *D).	Set no.7.
NH *t instead of *s from POC *ns (PCP *D?,*d?).	Set no.9.
NH *d instead of *s from POC *ns (PCP *D?,*d?).	Set no.10.

Other considerations which lead to the positing of the CP languages as a possible source will be mentioned in the discussion of the individual cases in the next section.

While the sets of this group point to the CP languages only as a possible source, those of the second group restrict the choice to the CP group since the latter could have borrowed the CP members of the sets from a NH language only, as was already argued in the introduction. The alternative, that the CP forms in these sets are lexical innovations of the CP group which were borrowed by the NH language(s) with which they were in contact is rejected on the following grounds:

1. The reconstructed NH forms in sets 13-18 all seem to have a final consonant¹³ which does not appear in the (reconstructed) CP forms. The simplest explanation is that the CP languages dropped those final consonants because they do not allow consonants in word-final position — the alternative being that the NH languages somehow acquired a number of consonantal accretions. The assumption is supported by the evidence in set 13, which will be discussed in full in the next section;

2. In two sets (nos 19,20) the CP forms display sound correspondences which do not fit into the pattern of reflexes of POC consonants and which mark them as of probable NAN origin;
3. Sets 21-26 are much more fully represented in the NH than in the CP group.¹⁴ The same argument, but then the other way around, applies to sets 27 and 28: in these the CP languages are more fully represented than the NH languages, which could be interpreted as a sign of borrowing by a NH language from a CP language.

The presentation of the evidence, given in the next section, is as follows: for each item the NH forms are given on the left, the CP forms on the right. Where the NH data allow the reconstruction of an earlier form this form is given preceded by an asterisk.¹⁵ The NH and CP data are followed where possible by the reconstructed forms of the CP words — in case of AN words, from PAN through POC down to the lowest level so far known to me. The data are generally followed by a short discussion. For the names of languages and language groups the following abbreviations have been used:

NH	North Halmahera(n)	CP	Central Papua(n)
PA	Pagu	MTU	Motu
MO	Modole	MEK	Mekeo
TO	Tobelo	ROR	Roro
LO	Loda	LAL	Lala
TB	Tobaru	KUN	Kuni
SA	Sahu	GAB	Gabadi
IB	Ibu (dialect of SA)	DOU	Doura
TE	Ternate	SIN	Sinagoro
TI	Tidore	HUL	Hula
WM	West Makian	KEA	Keapara
		ARM	Aroma

Proto-languages:

PAN	Proto-Austronesian	PML	Proto-Malaitan
PPH	Proto-Philippine	PSU	Proto-Suau
PAM	Proto-Ambonese	PCP	Proto-Central Papuan
POC	Proto-Oceanic	PWC	Proto-West-Central Papuan
PEO	Proto-Eastern Oceanic	PEC	Proto-East-Central Papuan
PPN	Proto-Polynesian	PMT	Proto-Motuic.

5. SETS OF PROBABLE COGNATES

5.1. Sets of Austronesian origin

1. 'arm, hand'

NH:	*giam(a) ¹⁶	CP:	
PA, MO	giam	MKE, ROR, KUN, GAB, DOU, MTU	ima
LO, TB, SA	giama	SIN	gima
GA, TE, TI	gia	HUL	gima
WM	ia	KEA	ima
		ARM	hima

PAN *lima', POC *lima, PCP *qima, PEC *gima, PMT *ima

The initial q in the PCP form is a true accretion (Lynch 1978). A tentative explanation of the medial a in the NH forms is that gima > giam and that this form subsequently lost final m or received a supporting vowel in those NH languages which developed a preference for vocalic word endings. The gima > giam type of metathesis is found in quite a few Oceanic languages¹⁷ and is also common in some of the AN languages spoken on the Aru islands. In the NH languages it is the only case I have come across.

2. 'branch, finger'

NH:	*ra(ga)-raga	CP:	
LO, GA	raraga 'fingers, branch'	MTU	raga 'branch of a palm bearing the fruit'
PA	raragar 'fingers'		
TE, TI	raga-raga 'fingers'	rege	'limb, tributary'
	rage-rage 'twigs'	rigirigi	'small branch'
SA	raraga, raga-raga 'fingers, toes'	HUL	ra
		KEA	raa
		ARM	raga
		SIN	lega

PAN *(d)ahan, POC *taqa(n), PEO *da'a 'branch', PPN *laqa 'small branch', PAM *lagat 'finger', PCP raqa, reqa 'branch'.

The source language must have had a form with initial r. Initial l remained unchanged, yielding laga. Note that PCP New Guinea area one often finds words meaning both 'branch' and 'finger'. In many Papuan languages the extremities of trees are equated with the extremities of trees.

3. 'good'

NH:	*loha	CP:	
PA, LO	loa	MTU	lohia
TO, TB	oa	KUN	lobia
TE, TI	laha		
SA	laa		
GA	loha		

PMT *lovia (?)

The MTU and KUN forms allow us to reconstruct PMT *lovia 'good'. NH *h : PMT (PWC, PEC, PCP) 'pointer' to the CP group as a possible source (see s. 1.1.1). This is found in one more case: 'warm', no.11.

4. 'lime'

NH:	*gahu	CP:	
GA	gahu	MTU, DOU	ahu
MO, LO	gau	KUN, ROR	abu
WM	afo	LAL	avu
		MEK	apu
		SIN	gagu
		ARM	gavu

PAN *'apuRe, PPH *qapur, PAM *apur, POC *qapuR, PCP *qavu, PEC *gavu, PMT *avu, PML *gau.

NH *h : CP *v and loss of POC *r both point to CP as a possible source. If the NH form originally had medial p and final r we would have found GA gapu, MO, TO gawuru, WM apo.

5. 'mouth, lips'

NH: *uru		CP:	
PA, MO	ulu	LAL	nutu
GA, LO, TO, TB	uru	DOU	utu
SA	u'du	MTU	udu
		SIN, HUL, KEA, ARM	murū

PAN *ŋu(ts)u, POC *ŋusu, PPN *ŋutu, PCP *ŋudu, PWC *NuDu, PMT *uDu, PEC *muru, PAM *ŋidu

Identifying feature: loss of initial nasal. This would point to a subgroup of PCP as a possible source (i.e. PMT). The source cannot have contained medial t or s which would have resulted in different series of correspondences in the NH languages. The PAM subgroup is ruled out because of the initial nasal and the high vowel in the proto-form.

6. 'to hide'

NH: *ihun(i)		CP:	
GA	ihu	MTU	huni
TO, MO, TB	iunu		

PAN *buni(qh'), POC *puni, PAM *vuni, PML *thaifuni

Data in other CP languages than Motu are lacking and no PCP form has yet been reconstructed. MTU huni however would be regularly derivable from POC *puni via a PCP form *vuni. In this case, PAM or one of its daughter languages could also have been the source of the NH form but in the light of the other evidence I think CP is a more probable source. The initial i in the NH forms could be due to metathesis (a case similar to NH *urit, PCP *turi, no.12) or the source language had a form with preposed i (cf. PML *tha-ifuni) but no evidence for such a prefix is available in my CP data.

7. 'to rub, wipe'

NH: *Dahusu		CP:	
GA	dahu	DOU	tahu
TO	ɗauhu	MTU	dahu
MO	dauhu	SIN	da
TB	dausu		

PAN *sapu, POC ?, PCP *Davu

I have not come across a POC reconstruction, but the expected form would be *sapu. The SIN form is irregular; one would expect ravu. The extra syllable in the NH form still calls for an explanation. A possibility, it seems to me, is that the NH languages have combined two semantically related AN roots into one compound stem, the second root being usu, from PAN *gusuk, POC *gusu(t). The case would be parallel to that of 'to tie', no.20.

8. 'smell'

NH: *bouna		CP:	
TO	bouna	MTU, SIN, ARM	bona
TB, SA	bounu		
TE	bou		

PAN *bahu, POC *ŋpona (?), PCP *bona

The TO form suggests that the form borrowed was bouna, and that in TB and SA the final vowel assimilated to the preceding u. The POC form *ŋpona was reconstructed

on the basis of PCP *bona* and Proto-North New Hebridean *kwona* (Ross 1979). But in the Loyalty islands the form *bun* is found¹⁹ and I would not exclude the possibility that the POC and PCP forms contained an *u*: **ḡpouna*, **bouna*.

9. '*companion, spouse*'

NH: *-ata (* <i>gatawa</i> ?)	CP:	
PA <i>wek-at</i> ' <i>wife</i> ' (<i>ḡoweka</i> ' <i>woman</i> ')	MEK	<i>akawa</i>
LO <i>lok-at</i> ' <i>husband</i> '	ROR	<i>atawa</i>
TO <i>hek-ata</i> ' <i>wife</i> ' (<i>oheka</i> ' <i>woman</i> ')	KUN, LAL	<i>ada</i>
LO <i>rok-ata</i> ' <i>husband</i> '	GAB	<i>agava</i>
MO <i>o'-ata</i> ' <i>husband</i> '	MTU	<i>adava</i>
WM <i>at</i>	HUL	<i>arawa</i>
	KEA	<i>harawa</i>
	ARM	<i>ḡarawa</i>

PAN **qasawa*', PPH **qasawa*, PAM **sawa*, POC **qansawa*, PCP **qaDawa* ?, PWC **adawa*, PMT **adawa*, PEC **ḡarawa*.

Identifying feature: the source language should have had a form with medial *d* or *t*.

In NH the AN root has been compounded with apparently NAN roots meaning '*woman/female*' and '*man/male*'. Intervocalic *w* is not well attested in the NH languages; PNH **w* seems to have been dropped generally and *ata* could well reflect an earlier **atawa*. The form borrowed by the NH languages could also have had an initial *g* which was lost after the velar stop of the preceding morpheme (NH languages did not allow word-medial consonant clusters), and regularly dropped in WM. The CP languages present a problem in that the West-Central forms point to a PCP form **qadawa* and a POC form **qantawa*, whereas the East-Central forms allow one to reconstruct PCP **qaDawa*, a regular descendant of POC **qansawa*. The PMT form could be a descendant of either PCP **qadawa* or **qaDawa*. Perhaps this is a case of doublets on the POC or PCP level.

10. '*above, upwards*'

NH: * <i>dake</i> , * <i>daku</i>	CP:	
GA, TO <i>dake</i> , <i>daku</i> ' <i>above</i> ' (near/far)	MTU	<i>dae</i>
SA <i>da'u</i> ' <i>above</i> '	MEK	<i>kae</i>
TE <i>daku</i> ' <i>above</i> '	ROR	<i>tae</i>
	LAL	<i>da'e</i>
	SIN	<i>rage</i>
	ARM	<i>ve/rage</i>

PAN **sakay*, PAM **saka*(y), POC **nsake*, PPN **hake*, PCP **dage*, **Dage*?, PWC, PMT **dae*, PEC **rage*.

The NH forms could only have come from a source with initial *d* or a phonetically similar stop. Note that the CP forms present the same problem as in the previous case: the PWC form reflects PCP **dage*, the PEC form reflects PCP **Dage* and the PMT form could reflect either of the two.

11. 'warm, hot'

NH: *sahuk(u)	CP:	
PA sasauk	MEK	iapu
GA, TE sahu	KUN	siabu
LO sau	LAL	siavuli
TO hauku	GAB	siau
MO, SA sau'u	DOU	siauri
WM sasafu	MTU	siahu
	HUL, KEA, ARM	iabu

PCP, PMT, PWC *tiavu

Ray (1907) links the CP forms with Banks Islands saw, sawsaw, sewsew. *tiavu therefore does not seem to be an PCP innovation, but so far no POC form has been reconstructed. PCP *t changed to s in the western and central CP languages when the following vowel was i; the NH forms suggest that this change had already taken place when the form was borrowed. CP ia : NH a in one more case ('good', no.3). The final k in the NH form remains unexplained.

12. 'to sew, to stitch'

NH: *urit(i)	CP:	
PA ulit	MTU	turi
GA uri	SIN	tuli
MO uliti	HUL, ARM	uli
TO uriti		

PCP *turi, PEC *tuli; PPN *tui

The NH forms would show metathesis of the originally final t to word-initial position. A similar case would be that of *huni > ihun, no.6.

5.2. Sets, probably of non-Austronesian origin

13. 'water, river'

NH: *akel(e)	CP:	
PA akel	MEK	ake
GA, TE, TI ake	ROR, LAL	ate
TO, LO, TB akere	GAB	aqe
MO a'ele		

PWC *ade ?

Ray (1907) links the CP forms with Wano (San Cristobal) ahe 'creek'; in the Shortland Islands the form atele is found.²⁰ It shows that the CP forms lost the last syllable, which was retained in several of the NH languages. Cognates are found as far west as north Sulawesi and the Sangir-Talaud islands (Bentena ake, Tonsawang axe, Sangir ake) but in those areas they probably reflect the political influence of the sultanate of Ternate. Probable cognates are found in many Papuan languages on the New Guinea mainland and they could very well be of non-Austronesian origin. The forms ake, ate in the western CP languages point to borrowing from the NH languages rather than from any Papuan source on New Guinea. The fact that related forms are also found in the southern Solomons will not concern us here. It is one of a few indications that the line of traffic could have extended east beyond the tip of Papua, a possibility which still awaits further research.

14. 'belly'

NH: *pokol(o)	CP:	
PA pokol	KUN	foa
GA poko	LAL	bo'a
LO, TB pokoro	DOU	boa
MO po'o	MTU, SIN	boga
SA po'olo	HUL	poka

PCP *boga, PSU *boga

PAM has *vuka; its relationship to the forms above is unclear. Note the parallel with the previous item: again the NH forms contain a final consonant which is missing in the CP forms. The same is found in items 15-18.

15. 'canoe'

NH: *(η)otil	CP:	
PA, SA, TE, TI oti	ROR	ahi
TO, TB notiri	LAL, DOU, MTU	asi
MO ɣotili	GAB	asi ([atsi])
WM eti	SIN	ɣasi, ɣati
	HUL	ai

PCP *qati, PEC *ɣati, PMT *ati

Ross (1979) links PCP *qati with PAN *katiR 'outrigger boom' and reconstructs POC *katiR on the basis of the PCP reconstruction. If the NH form *ɣotili is an AN loan, the source language cannot have had initial k in this word; nor could it have been PCP *qati which would yield gati in NH. However, the validity of the PCP reconstruction is not beyond doubt; the Motu form should then have been ɣasi, not asi. Equally possible is that the PCP reconstruction should be *ɣati which would make the MTU form regular, but the HUL form irregular (ai instead of ɣai). If the form ɣotili was borrowed from a NH language at the PCP stage, -l- (between two high vowels) would have been dropped, and the development would be ɣotili > ɣotii > ɣoti > ɣati, etc.

Another possibility is that the initial nasal in the NH forms is an accretion. There are more cases of such a nasal accretion in NH words, a phenomenon for which no explanation has yet been found.²¹

It is interesting that the Eleman languages further to the west have possibly related forms oroti, loti: see section 7.

16. 'cloud, fog, mist'

NH: *kamol(o)	CP:	
PA kamo-kamol	SIN, HUL	ɣamau 'mist'
TO, MO kamo-kamoro	ARM	amau
SA kamokamo		
TE, TI kamo		

PEC *ɣamau

Probable cognates of NH *kamol have so far only been noted in the East-Central subgroup of the CP languages. The situation is not clear: borrowing from PEC should have resulted in NH forms with initial g; borrowing from a NH language in CP forms with initial k.

17. 'head'

NH: *dopolo, *depolo ?	CP:
TE dopolo	SIN deba, tepa
TI dofolo	HUL reba, leba
PEC *deba	

There are several instances known of TE and TI changing *e to o: NH *besaka 'rain' > boha, *pereto 'narrow' > poreto, **gekono* 'path' > *goko*, amongst others. At an earlier stage the NH form could have been *depolo, but support from the other NH languages is lacking, as they have unrelated forms. Note the parallels in the sound correspondences between *depolo and *deba, *pokolo and *boga ('belly', no.14).

18. 'sky, dark cloud, black'

NH: *dipaN(a), *tupam	CP:
PA liwan, diwan	MTU duba 'very dark cloud, very dark colour'
GA dipa	EC languages dubara 'black'
LO diwama	ruparupa
TO dihaŋa	rubaruba
MO, TB, SA 'diwaŋa	tupatupa
TE, TI tufa	
WM tupam	

The whole NH set is irregular in that one would expect TE, TI, WM to have difa and dipam, or the other languages to have forms with initial t followed by u. Perhaps it points to a doublet *dipam, *tupam in PNH. The latter could be the source of the CP forms. The semantic shift between the two extremes 'sky' and 'black' (sky > blue sky > dark cloud > dark colour, black) is not uncommon in the NAN languages I am acquainted with and in the CP languages we find descendants of PCP *guba meaning both 'sky' and 'dark cloud'.

19. 'to bite'

NH: *goli	CP:
GA, TO, TB goli	ROR 'uri
SA go'di	MTU kori
TE logi	SIN koli
	KEA, ARM oli

PCP *kori, PEC *koli, PMT *kori

Consonant metathesis is common in Ternate. The *kori series seems to be restricted to the CP languages and is found beside regular descendants of PAN *kaRat 'to bite': PCP *qara, PMT *ara, PEC *gala. The initial k in *kori marks the form as one of probably non-Austronesian origin, and a NH language seems a plausible source.

20. 'to tie'

NH: *pili-kutu ?	CP:
PA, GA piliku	MTU kwatu 'to tie a knot'
LO pilikutu	
TO dikutu	
MO likitu	
SA piri?u	
TE podiku	

PCP ?, PML *gwathu

The NH forms seem to contain a common AN root *pili* (PAN **pi(dDr)i(Ct)* 'to plait' POC **pi(dr)i*, PEO **piri*, PML **firi*; MTU *hiri* 'to fasten by twisting round and round'). This root has been compounded with a semantically related root *kutu* which has probable cognates in Motu and in PML. Words with initial *kw-* in Motu are also suspect of being of non-Austronesian origin. The source again could be a NH language; in NH the original non-Austronesian root would then at some stage have been combined with a newly introduced, semantically related AN loan.

21. 'middle, centre'

NH: * <i>golona</i>	CP:
TO <i>gorona</i>	MTU <i>dina gelona</i> 'midday'
SA <i>golona</i>	
TB <i>wane i korona</i> 'midday'	

TB *wane*, MTU *dina* = 'sun'. TB *korona* is the verb ('to be in the middle') corresponding to the noun *gorona*. Data in other CP languages are lacking.

22. 'to climb'

NH: * <i>dola</i>	CP:
PA, GA, LO <i>dola</i>	MTU <i>dara</i>
TO, MO, TB <i>doa</i>	

Motu has also another verb 'ascend, climb': *daekau* which contains a morpheme *dae* 'upwards', another possible link with NH. See no.10.

23. 'cold(ness)'

NH: * <i>alo</i>	CP:
GA, LO, SA, TE, TI <i>alo, alo-alo</i>	KUN <i>aulolo</i> 'cold'
	'coldness'

The Kuni form can be found in Ray 1907 given as a dialectal form collected by Kowald (1894). Another Kuni form is *elu*, from PCP **keRuma*, not a cognate of NH *alo*.

24. 'to hear, to know'

NH: * <i>isen(e)</i>	CP:
PA <i>isen</i>	GAB <i>isana</i> 'to know'
GA, TE <i>ise</i>	MTU <i>iseuna tauna</i> 'expert person'
TO, MO <i>ihene</i>	
TB, SA <i>isene</i>	

In preliterate societies in which people learn by listening, not by reading, the concepts of hearing and knowing are intimately connected and often form semantic aspects of one and the same word. Or, if 'to hear' and 'to know' are different verbs they contain the same root. Ray (1907) for example gives MEK *kono* 'to hear, know'; LAL *ika* 'to hear', *ika-basi* 'to know'.

25. 'to stand up'

NH: * <i>momik(i)</i>	CP:
PA <i>momik</i>	ROR <i>mikiri</i>
TE, GA <i>momi</i>	
LO, TO, TB <i>momiki</i>	
MO, SA <i>momi'i</i>	

All NH forms contain the root *mik(i)* which has been reduplicated by a normal process in which the duplicate syllable receives the vowel *o*. ROR *-ri* remains unexplained.

26. 'voice'

NH: *elen(i) or *elin(i)
 GA ili
 TE oli
 TI idi
 SA i'diqi, eliqi 'remember'
 LO, TB iliqi
 TO elenji

CP:
 MTU ere 'voice, speech, word'

5.3. Residual sets

27. 'to put down, plant'

NH:
 GA hado
 TO fato

CP:
 ROR bato
 KUN bado
 LAL vado
 MTU hado
 SIN, ARM varo
 HUL vari

PCP, PWC, PMT *vado, PEC *varo

The TO form is in the Boeng dialect which has f where Tobelo proper has h but the corresponding sound in GA should be p, not h.

28. 'to say, speak'

NH:
 TO ato 'to say'

CP:
 MEK aiso 'speech, voice, throat'
 ROR ako
 KUN ajo
 LAL, DOU ato
 GAB ago
 MTU gado
 SIN, HUL garo

PCP *gaDo, PWC *qaDo, PMT *gaDo, PEC *garo

Thus far I have found a NH cognate only in Tobelo; note the parallel with TO ole 'tail bone' (see no.29) in which word TO also seems to have lost an initial g. Also noteworthy is that TO, TB have tomara 'throat', and MTU toma 'to say, think', a set of probable cognates in which the semantic relation is the reverse of the one in the ato - gaDo set.

29. 'buttocks'

NH: *gola(la)
 LO gola
 MO goa-gono
 IB golala
 TO ole 'tail bone'

CP:
 HUL, KEA kore, gore, kole, gare
 MTU gole 'tail of bird'

PEC *gore?, *gole?

30. 'face'

NH: *bira
GA bira

CP:
ARM bira

ARM would reflect PEC *bira. Most other NH languages have biono, and also ARM bira seems to be an isolated form.

31. 'to play'

NH: *ule
GA, LO, TO ule

CP:
HUL, KEA, ARM ula-ula

6. NH - CP CONTACT: HOW AND WHERE

If the 31 sets of probable cognates indeed, as I interpret them, reflect contact between speakers of languages ancestral to those spoken in the North Moluccas and Central Papua, the question arises where this contact took place and how it came about. As for the first part of the question the historical evidence favours the Central Papua area as the place of contact. The Motuans who covered the westernmost part of the network of trade relations along the coast of south-east New Guinea do not seem to have ventured further west than the delta of the Purari river in the Gulf of Papua.²² The natives of the North Moluccas or 'Alfuren' as they were called enter our historical records as highly skilled seafarers whose expeditions took them as far south as the Lesser Sunda Islands and as far east as Etna Bay on the south coast of Irian Jaya. These skills which they probably acquired during early contacts with Austronesians moving through the Moluccan/West New Guinean area²³ were quite sufficient to take them to the eastern tip of Papua if they wished. Although there are no historical records of such voyages - and here we come to the second part of the question - there is some evidence from in-between points which suggests that they indeed took place and that they roughly followed the south coast of New Guinea. The evidence unearthed so far comes from three points: the Eleman languages, the Kiwai languages, and the eastern part of the Aru Islands.

The Eleman languages are found along the coast of the Papuan Gulf immediately west of the CP languages. The two most important of them are Toariipi (TOA) and Orokelo (ORO). These have borrowed quite a few words from their eastern (CP) neighbours as can be expected, but they also contain a few AN words which probably came from the west and, more important, some NAN words which have probable cognates in the NH languages:

AN items:

'to break'	TOA topuka- ²⁴	NH *tobik(i)	PAN *tebik
'dry'	TOA arara ORO kakarara	(NH *kokasa)	PAN *karaN

Probable cognates of the Eleman forms are found along the south coast of the Bird's Head and in the Bomberai Peninsula: Puragi kaker, Negeri Besar kara, Arandai garagare, Karas karara, Iria tatara. All these languages are non-Austronesian and the presence of such similar AN loans in them I interpret as a sign that there was AN traffic from west to east all along the south coast and that the Alfuresse ancestors only followed an already existing (trade?) route. The NH form comes from an Oceanic source, cf. POC *kasan (for the reduplication see section 5. no.25).

'frog' TOA pakeke NH *(po)padeke PPH *paNkaq

NAN items:

'to carry' TOA taio-²⁵ NH *Tahi
 'muddy' TOA pekauke NH *(pe)peke
 'star' TOA koru NH *koruk
 'to steal' TOA torea loi NH *torik
 'thunder' TOA tutururu NH *duturu
 'canoe' TOA oroti NH *(ŋ)otil(i)

This case has already been mentioned in section 5. no.15.

The Kiwai language, spoken on Kiwai island in the delta of the Fly River, offers much the same picture:

AN items:

'father' KIW abera, baba NH *aba, *baba PAN *bapah
 'rope, ratten' KIW iwi NH *iwi PAN *'uvay, POC *quwe
 'this' KIW ni NH *ne-ena
 'that' KIW gi NH *ge-ena

NAN items:

'to eat, to drink' KIW odio ('drink') NH *oDom(o) ('eat')
 'sibling' KIW nira dubu NH *nyira
 KIW dubu = 'man';
 nira dubu = '2nd brother';
 TE nyira = 'elder brother'
 'elder sister' KIW abida NH *biraŋ(a)
 'to sleep, lie down' KIW utu ('sleep') NH *idu ('lie down')
 'fin of fish' KIW aibi NH *aebe
 'tree' KIW ota NH *gota

The proto-Kiwai form is *kota.

The Kiwai and Eleman languages are members of the Trans-New Guinea Phylum, the NH languages of the West Papuan Phylum. The chance that any of the NAN cognates shared by these languages and the NH languages reflects true genetic relationship is therefore very small, although the possibility cannot be excluded.

My data in the languages AN of the Aru Islands are very restricted; those quoted below are from Banjaring village on the east coast. Of the three forms two seem to have exclusively shared cognates in Motu, and one has cognates only in the NH languages:

'to cook' ARU nai MTU nanai 'warm the hands over fire'
 'dry' ARU marmarei MTU marai
 'back of body' ARU tudin NH *dudun(u)

The first two sets could be an indication that in prehistoric times the Motu travelled much farther west than they did when they were first contacted by Europeans.

7. SOME NOTES ON THE SOURCES CONSULTED

Almost all serious descriptive and comparative work in the Halmaheran languages has been carried out by Dutch missionaries of the Utrechtse Zendings-Vereeniging who worked in the area between 1866 and 1925: M.J. van Baarda, H. van Dijken, G.J. Ellen, J. Fortgens, A. Hueting, and G. Maan.²⁶ To them can be added the missionaries Adriani and Kruyt who worked among the Toraja of central Sulawesi. In the first edition of their great three-volume work on the Bare'e-speaking Toraja they included a comparative study of the AN languages of south Halmahera. The most important recent work is the sociolinguistic study by E.K.M. Masinambow. Linguistic data have further been collected in recent years by a few anthropologists working in Halmahera. I am indebted to two of them, Michael Young (Australian National University) and Paul Taylor (Smithsonian Institution) for providing me with wordlists collected by them.²⁷

The Central Papuan languages have attracted a lot of attention in recent years. As a result several descriptive and comparative studies have become available although not all of them have been published. For this paper I relied mainly on the work of A.K. Pawley, J.D. Lynch, M. Ross, and T.E. Dutton. The latter kindly gave me access to his own unpublished lexical data in the East-Central languages. Of the older works I used R. Lister-Turner and J.B. Clark's Motu dictionary, and the comparative lists of the Melanesian languages in Papua compiled by S.H. Ray.

For the Eleman languages I consulted H.A. Brown's Toaripi dictionary and Ray's old wordlists. The only source for Kiwai is still Ray's Kiwai grammar which contains a vocabulary compiled by E. Baxter Riley. The Aru wordlist from Banjaring village was kindly made available to me by Professor J.J. Fox of the Australian National University.

The data in the languages of the Bird's Head are from my own collection of wordlists. For the languages of Ambon and Seram I consulted Stresemann's *Lauterscheiningen in den ambonischen Sprachen*. Finally, the whole exercise would not have been possible without the Proto-Austronesian finderlist of S.A. Wurm and B. Wilson.

8. NOTES

1. This is a completely revised version of the paper originally presented to TICAL.
2. For more detailed information on these two phyla the reader is referred to the surveys in Wurm, ed. 1975, section 2.5. - 2.10.
3. The most recent assessment of their genetic relationship is given in Voorhoeve forthcoming.
4. A stock-level family is a language family which is the only member of a language stock. In lexicostatistical terms it means that its languages share from 28% - 81% cognates between them but less than 12% with the most closely related languages outside the family. The NH languages have been surveyed in more detail in Voorhoeve forthcoming and 1982. For their location see the map, p.218.

5. They are: Tehit, Kalabra, Seget, Moi, and Moraid (West Bird's Head stock-level Family); Madik and Karon Pantai (North Bird's Head Family) and Amberbaken, a stock-level isolate. More details are given in Wurm, ed. 1975, 2.10.; for their location see the map, p.218.
6. See note 3.
7. Maan 1951 gives the numbers of speakers for six of them: Buli 1000, Maba 1800, Patani 2700, Sawai 800, Weda 900, Gimán (Gane) 1500. This compares with Ternate 42,000, Tidore 26,000 (Masinambow 1972), West Makian 12,000 (Voorhoeve 1982), Sahu 12,000, Tobaru 10,000, Loda 13,000, Galela 15,000, Tobelo 15,000, Modole 1800, Pagu 2000 (Voorhoeve forthcoming).
8. Blust 1978.
9. Dutton (1976:620) suggested that they have their closest links with Sinagoro, especially its western dialects. Lynch (1978) however suggested that the Magori subgroup belongs with the West-Central languages. And Ross (1979a) after a reassessment of the data leaves the matter undecided.
10. POC = Proto-Oceanic.
11. As a systematic reconstruction of the Proto-North Halmaheran (PNH) sounds and lexicon is still in its initial stage, the asterisk is here used to mark the earliest form of a sound or word reconstructable on the basis of the available data. For loanwords this means the form in which it was supposedly borrowed.
12. The expectation is based on the PAN, POC etc. reflexes most frequently found in AN loans in the NH languages. It would exceed the limits set for this paper to present this material here and I have to ask the reader to take my word for it.
13. I say "seem to" because, although there are arguments for PNH allowing word-final consonants, the matter is still undecided. For this reason lexical reconstructions which probably ended in a consonant have a bracketed final vowel added to it (section 5, *passim*).
14. Admittedly a weak argument, as the uneven representation may reflect gaps in the data.
15. See note 11.
16. See note 13.
17. D.T. Tryon, personal communication.
18. See the table of CP sound correspondences, Appendix part 3.
19. D.T. Tryon, personal communication.
20. D.T. Tryon, personal communication.
21. See the short discussion of this phenomenon in Voorhoeve 1982, chapter 4.
22. The most recent publication dealing with various aspects of the so-called Hiri trade expeditions is Dutton, ed. 1982.

23. Milke (1961:180) proposed the hypothesis that the speakers of POC lived for some time in the proximity of the languages ancestral to those now spoken in the Philippines, Sulawesi, and East Indonesia. If he is right it means that these early contacts go back for at least two thousand years.
24. The dictionary form is *topukavai*.
25. The dictionary forms is *taiovai*.
26. Several of their publications appeared after 1925; the last one dating from 1951.
27. The data have since been supplemented by materials I collected in the north Moluccas shortly after the completion of the first version of this paper.

9. APPENDIX

9.1. Phoneme inventory of the North Halmaheran languages

The following is only a tentative inventory because only the West Makian and Sahu sound systems have so far been analysed in phonological terms; for the remaining languages we still have to rely on the old Dutch sources and my own impressions based on my fieldnotes. Phonemes which do not occur in all languages have been underlined and their distribution is given below. The symbols have their usual phonetic value except for *f* which is a bilabial fricative, *'b*, *'d*, *'j* which are implosive voiced stops and *ɖ* which is a voiced interdental fricative.

Consonants:

p	t	c	k	'
b	d	j	g	
<u>'b</u>	<u>'d</u>	<u>'j</u>	<u>'g</u>	
m	n	ny	ng	
<u>f</u>	s	<u>ɖ</u>		<u>h</u>
	l			
	r			
w	y			

Vowels:

i	ü	u
e	<u>ě</u>	o
	a	

' is found only in Modole and Sahu.

The implosive stops occur in Sahu only.

f not in Modole.

h not in Pagu.

ɖ only in Tobelo; phonemic status uncertain.

ü reported for Galela, Tobelo, and Modole only. Phonemic status uncertain.

ě [ə] as a phoneme only in Sahu.

9.2. Consonant correspondences between the North Halmaheran languages

The correspondences tabulated below have been tentatively worked out on the basis of the available data.

NH	*p	*P	*b	*m	*w	*t	*d	*n	*s	*r	*l
PA	p	w	b	m	w	t	d	n	s	l	l
MO	p	w	b	m	w, ϕ^3	t	d	n	s, h ²	l, ϕ^4	l, ϕ^4
TO	p	h, f ²	b	m	w	t	d	n	h	r	l, ϕ^4
GA	p	p	b	m	w	t	d	n	s	r, ϕ^5	l
LO	p	w	b	m	w	t	d	n	s	r	l
TB	p	p, w ³	b	m	ϕ	t	d	n	s	r	l, ϕ^4
SA	p	w	b, 'b ⁴	m	w, ϕ^3	t	d	n	s	r, 'd, ϕ^5	l, 'd, 's ⁵
TE	p, f ¹	f	b	m	w	t	d	n	s	r, d, ϕ^5	l, d ⁵
TI	?	f	b	m	?	t	d, l ¹	n	s	r, ϕ^5	l
WM	?	p	b	m	?	t	d	n	s	?	l

NH	*c	*j	*ny	*k	*g	*ng	*h	*D	*S
PA	t	d	n	k	g	ng	ϕ	y	l
MO	t	d	n	'	g	ng	ϕ	d, ϕ^6	h
TO	t	d	n	k	g	ng	h, ϕ^4	d	h
GA	t	d	n	k	g	ng	h	d	s
LO	t	d	n	k	g	ng	h, ϕ^4	j	j, r ¹
TB	t	d	n	k	g	ng	ϕ	d, ϕ^6	s
SA	c	j	ny	'	g, k, ϕ^1	ng	ϕ	r	r
TE	c	j	ny	k	g, k, ϕ^1	ϕ	h	h	r
TI	c	j	?	k	g	ϕ	h	y	?
WM	j	?	?	ϕ, k^3	ϕ, g^3	ϕ, ng^3	f	ϕ	?

¹No conditioning factor evident.

²Dialectal difference.

³The first reflex occurs in initial position, the second in medial position.

⁴Phonologically conditioned.

⁵Conditioned by the position in the word and by phonological factors.

⁶d in non-verbs, ϕ in verbs.

9.3. Summary of the major consonant correspondences between POC, PCP, and the Central Papuan languages according to Ross 1979

POC	-	p	mp	m	w	t	nt, nj	ns, s
PCP	p	v	b	m	w	t	d	D
PMT	p	v	b	m	w	t	d	D
MTU	p	h	b	m	v	t (s) ⁵	d	d
PWC	p	v	b	m	w	T	d	D
LAL	p, ϕ	v	b	m	v	k (s) ⁶	d (t) ⁹	t
KUN	f	b	f	m	v	k (s) ⁶	d	k (s) ⁷
GAB		v (ϕ) ¹	b	m	v	k (s) ⁶	d (g) ⁵	g
DOU		h	b	m	v	k (s) ⁶	t	t
ROR		b (h) ²	p	m	b, w ⁴	h	t	k
MEK	p	p (ϕ) ³	f	m	f	'	k	'

PEC	p	v	b	m	w	t	d	r	
ARM	p	v	b	m	w	∅	r	r	
HUL	p	v	b	m	w	∅	r	r	
SIN	p	v (ɘ) ⁸	b	m	w	t	d (r) ⁹	r	
POC	d,nd,r	l,y	n,ñ	ŋ	-	ŋk	k,q	-	
PCP	r	R	n	ŋ	k	g	q	kw	
PMT	r	R	n	N	k	g	ɘ	kw	gw
MTU	r	l	n	∅	k	g	ɘ	kw	gw
PWC	r	R	n	N	q	∅	∅	qw	
LAL	l	l	n	n,∅ ⁹	∅	∅	∅	∅	
KUN	l	j	n	n,∅ ⁹	∅	∅	∅	∅	
GAB	r	∅	n	n,∅ ⁹	∅	∅	∅	∅	
DOU	r	l	n	n,∅ ⁹	∅	∅	∅	∅	
ROR	r	∅ (e) ⁹	n	n,∅ ⁹	∅	∅	∅	∅ ¹ ,∅	
MEK	ŋ	∅, l ⁹	n	n,∅ ⁹	∅	∅	∅	∅	
PEC	l	R	n	ŋ	kk	g	ɘ	kw	?
ARM	l	d	n	ɘ	∅∅	k	ɘ,∅ ⁹	w	
HUL	l	∅	n	ɘ	kk	k	ɘ,∅ ⁹	kw	
SIN	l	∅	n	ɘ	kk	g,ɘ ⁹	ɘ	kw	

¹PCP *v > ∅ / u²PCP *v sometimes > h / u³PCP *v sometimes > ∅ / u⁴PCP *w > w / a⁵PCP *t > s / i,e⁶PCP *t > s / i and in GAB also / i⁷PCP *D > s / i⁸PCP *v > ɘ before rounded vowels⁹No distributional rule evident

BIBLIOGRAPHY

BAARDA, M.J. van

- 1895 *Een Galelareaesch-Hollandsche woordenlijst.* The Hague: Nijhoff.
- 1904 *Het Loda'sch, in vergelijking met het Galela'sch dialect op Halmaheira. Bijdragen Koninklijk Instituut voor Taal-, Land- en Volkenkunde (BijdrTLV) 56:317-496.*
- 1908 *Leidraad bij het bestuderen van het Galela'sch dialect.* The Hague: Nijhoff.

BLUST, Robert A.

- 1978 *Eastern Malayo-Polynesian: a subgrouping argument.* In: Wurm and Carrington, eds 1978:181-234.

BROWN, H.A.

- 1968 *A dictionary of Toaripi with English-Toaripi index.* 2 vols. *Oceanic Linguistic Monographs* 11. Sydney: The University of Sydney.

DIJKEN, H. van, and M.J. van BAARDA

- 1895 *O Galêla-ka manga totoādé, manga tjarita déo manga pitūa. Fabelen, verhalen en overleveringen der Galêlareezen ... BijdrTLV 45:192-290, 387-560.*

DUTTON, T.E.

- 1976 *Magori and similar languages of south-east Papua.* In: Wurm, ed. 1976:581-636.
- 1982 *Borrowing in Austronesian and non-Austronesian languages of coastal south-east mainland Papua New Guinea.* In: Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.1: *Currents in Oceanic.* PL, C-74:109-177.

DUTTON, T.E., ed.

- 1975 *Studies in languages of central and south-east Papua.* PL, C-29.
- 1982 *The Hiri in history: further aspects of long distance Motu trade in Central Papua.* *Pacific Research Monograph* 8. Canberra: Research School of Pacific Studies, Australian National University.

ELLEN, G.J.

- 1916a *Woordenlijst van het Modòle op Noord-Halmahera.* *BijdrTLV 72:103-139.*
- 1916b *Woordenlijst van het Pague op Noord-Halmahera.* *BijdrTLV 72:65-102.*

FORTGENS, J.

- 1917 *Woordenlijst van het Ternatesch.* Semarang: Van Dorp.
- 1928 *Grammatikale aantekeningen van het Tabaroesch, Tabaroesche volksverhalen en raadsels.* *BijdrTLV 84:300-544.*

HUETING, A.

- 1908a Iets over de 'Ternataansch-Halmaherasche' taalgroep. *BijdrTLV* 60:370-411.
- 1908b *Tobèloreesch-Hollandsch woordenboek*. The Hague: Nijhoff.
- 1935 Supplement op het Tobèloreesch woordenboek. *BijdrTLV* 92:161-176.

KOWALD, C.

- 1894 Vocabulary: Upper Angabunga or Arabule. *British New Guinea annual report for 1892-93*, 113-114. Brisbane: Government Printer.

LeBAR, F.M., ed.

- 1972 *Ethnic groups of insular Southeast Asia*, vol.1. New Haven: Human Relations Area Files Press.

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

- n.d. *A dictionary of the Motu language of Papua*. Second edition, edited [c.1954] by Percy Chatterton. Sydney: Government Printer.

LYNCH, J.D.

- 1978 Proto-Central Papuan: a reassessment. University of Papua New Guinea. TS.

MAAN, G.

- 1940 *Boelisch-Nederlandsche woordenlijst met Nederlandsch-Boelisch register*. *Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen* 74/3. Bandung: A.C. Nix.

MASINAMBOW, E.K.M.

- 1972 Halmahera. In: LeBar, ed. 1972:119-122.
- 1976 Konvergensi etnolinguistis di Halmahera Tengah, sebuah analisa pendahuluan. Doctoral thesis, University of Indonesia.

MILKE, Wilhelm

- 1961 Beiträge zur ozeanischen Linguistik. *Zeitschrift für Ethnologie* 86:162-182.

PAWLEY, A.K.

- 1975 The relationships of the Austronesian languages of Central Papua: a preliminary study. In: Dutton, ed. 1975:3-105.

RAY, S.H.

- 1907a *Reports of the Cambridge Anthropological Expedition to Torres Straits*, vol.3: *Linguistics*. Cambridge: Cambridge University Press.
- 1907b A grammar of the Kiwai language, with notes on the Mawata dialect. In: Ray 1907 Reports . . ., 302-319.

ROSS, Malcolm

- 1979 Reconstructing Proto-Central Papuan. University of Papua New Guinea/Goroka Teachers' College. TS.
- 1979 The Austronesian languages of Papua: towards a family tree. Goroka Teachers' College. Mimeo.

STRESEMANN, E.

- 1927 *Die Lauterscheinungen in den ambonischen Sprachen. Zeitschrift für Eingeborenen-Sprachen.*

VEEN, H. van der

- 1915 *De Noord-Halmahera'se taalgroep tegenover de Austronesiese talen.* Leiden: Van Niffterik.

VOORHOEVE, C.L.

- 1975 *Languages of Irian Jaya: checklist, preliminary classification, language maps, wordlists.* PL, B-31.
- 1982 The West Makian language. In: Voorhoeve, ed. 1982:1-74.
- f/c Sahu dictionary. MS.

VOORHOEVE, C.L., ed.

- 1982 *The Makian languages, with North Moluccan vocabularies.* PL, D-46.

WURM, S.A., ed.

- 1975 *New Guinea area languages and language study, vol.1: Papuan languages and the New Guinea linguistic scene.* PL, C-38.
- 1976 *New Guinea area languages and language study, vol.2: Austronesian languages.* PL, C-39.

WURM, S.A. and Lois CARRINGTON, eds

- 1978 *Second International Conference on Austronesian Linguistics: proceedings.* PL, C-61.

WURM, S.A. and B. WILSON

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

SOUND CHANGES IN THE GORONTALO LANGUAGE

J. Noorduyn

1. INTRODUCTION

The Gorontalo language (G) is spoken in the town of Gorontalo and its environs on the northernmost peninsula of the island of Sulawesi or Celebes (Indonesia), where the coasts of this long and narrow peninsula run in a generally west-east direction. The town of Gorontalo is located on the south coast.

G is the principal and best known dialect in a larger language area, which also includes other closely related languages or dialects, such as the Bunda dialect (Bd), which is spoken in the Suwawa district to the east of the town of Gorontalo.

The languages adjacent to the Gorontalo language area are those of Bolaang-Mongondow (BM) to the east, and Buol, i.e. Bwuolo (Bw), to the west.

Throughout its vocabulary as well as in many of its affixes G shows the effect of a comparatively large number of sound changes which must have taken place in the relatively recent history of the language. As a result the phonetic shape of many words has often radically changed and several morphophonemic alterations have come into being in its morphology.

Since no data from older stages of G have been preserved, these sound changes and the previous situation have to be reconstructed by comparing the present-day G data with closely related languages or dialects in which these sound changes have not or have only partially taken place.

The language which has been used for this comparative purpose is BM, because it has turned out that many BM words and affixes have the shape which their G cognates must have had prior to the G sound changes. The synchronic evidence emerging from the comparison consists of rules of systematic correspondence between the G - BM cognates. The diachronic hypothesis arising from these rules of correspondence is that sound changes corresponding to these rules have operated in G, starting from original forms which had the same shape as the BM cognates concerned.

From the particulars of these sound changes a further diachronic specification emerges in that these sound changes appear to have operated in a definite relative order, which to a large extent can be reconstructed.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 241-261. *Pacific Linguistics*, C-75, 1982.

© J. Noorduyn

241

It follows from the above that this G - BM comparison is a low-level one, contributing to recognising G and BM as belonging to a first-level grouping and to establishing a G-BM proto-language, in which no higher-level grouping or proto-language needs to play a role. But, even though some remarks as to the place of G-BM in a higher level grouping can be made, the next step should be to put the G and BM data in the wider view of their PAN ancestry. Another step should be to investigate whether there are other languages which belong to the same grouping as G and BM, one of the likely candidates being Bw. This is possible only when sufficient data from such other languages become available.

It has been possible to use BM for this comparative purpose because a large quantity of BM data is available in the grammar and dictionary published by Dunnebier (1929-30 and 1951).

The G data used are mainly those published in Badudu's morphology of the verb (1982) and Pateda's small dictionary (1977).

In principle the Bd dialect can also be profitably used in this context for comparison with G and BM cognates, but at present this is only possible to a certain extent because of the insufficiency of the available data. The main data and tentative conclusions concerning this dialect are therefore put together in a separate section. For Bd there are only available the G - Bd - Malay - Dutch wordlist published by Schröder in 1908, which contains quite a few printing errors and errors of notation, and a report on the structure of the Suwawa language by M.M. Kasim and others (1979), which does not include a list of lexical data.

A few lexical data of the Bw language are to be found in Van Andel's list of terms of customary law published in 1929. Some observations based on these data are given in a separate section.

2. THE PHONEMES OF G (INCLUDING Bd) AND BM

Both languages have the following phonemes, except for those in brackets, which occur in BM only.

Vowels: a, e, i, o, u

Diphthong: (ai), (au), (ao), (ea), (oi), (ui)

Consonants: p, b, m, mb, (mp)
 t, d, n, nd, (nt)
 c, j, ñ
 k, g, ŋ, ŋg, (ŋk), (ŋŋ)
 l, r, y, w, s, (ns)
 ?, h

1. G vowels occur either short or long, e.g. a or a:.
2. BM t preceding i is replaced by s.
3. G and BM palatals occur only in loanwords, BM ñ also in the 3rd person possessive suffix.
4. G has no final consonants; in BM all consonants except the prenasalised ones, the palatals, y and h may occur word finally.

3. A NOTE ON SPELLING

The spelling systems used for G by Badudu and by Pateda differ in one major respect. The former writes all glottal stops except word initially, using q, and leaves out w and y after u and i respectively. The latter writes all w and y wherever they occur but no glottal stop. For instance, the sequence ua means uwa in the former's spelling and u'a in the latter's. Both alternatives give a consistent spelling. For the sake of clarity we use a combination of these spelling systems, writing all glottal stops, even word initially, as well as all w and y.

Badudu defines nt as a voiced consonant (konsonan bersuara, 1982:16). Therefore we write nd in words in which Badudu and Pateda use nt. Probably the d in nd is slightly different phonetically from the intervocalic or initial d.

For BM we use Dunnebier's spelling, apart from such usual replacements as u instead of oe.

4. THE G SOUND CHANGES

The Gorontalo sound changes which have been found in the comparative manner outlined above are listed below, each one illustrated by an example of BM-G cognates, and in a sequence which is a highly random one as it is determined to a large degree by the particulars of the available examples. Since many examples show simultaneously the effect of more than one sound change, the examples in the list have for the sake of convenience been selected in such a way that none of them shows the effect of a sound change mentioned lower in the list. The sequence of the sound changes listed has been adapted to this principle wherever necessary. In this way several sound changes are illustrated more than once but, besides at their proper place, only in examples occurring lower in the list.

Some further particulars applying to this list are the following.

1. The term 'sound change' is for the sake of convenience defined as applying to loss or addition of a phoneme as well as to change of one phoneme into another.
2. The figures in brackets (1), (2), etc., which are used to indicate the successive sound changes in the list, serve as references to the sound changes as they occur in the list wherever they are mentioned again in this article.
3. Single consonants mentioned in formulating the sound changes are always meant as those occurring either word initially or intervocalically. Where prenasalised consonants are meant, these are mentioned separately as such.
4. The words 'before', 'after', and 'following' as used in formulating the sound changes are always meant to denote the position *immediately* before or after a particular phoneme.
5. The BM-G cognates selected as examples in the list are always basically each other's exact semantic equivalents in both languages, except in two cases. In BM another word than anak (sub (6)) is in use for 'child', but there exist the derivations mojanak meaning 'young, small (of children)' and inanakan 'large family, clan'. The second example mentioned sub (13) is the only example found of the change of -mbo- to -mu-.

6. List of sound changes:
(in arbitrary order)

	BM	G	Examples:
(1) final ʔ was lost	puluʔ	pulu	'ten'
(2) final a became o	mata	mato	'eye'
(3) k became ʔ	kita	ʔito	'we'
(4) initial ʔ before i became y and before other vowels w	iko ata	yiʔo wato	'you' 'slave'
(5) o was added after final consonant	utat	wutato	'brother'
(6) n became l	-anak	walaʔo	'child'
(7) g became h	dugi	duhi	'thorn'
(8) b before u became h	bulan	hulalo	'moon'
(9) r became l	ruit	luwito	'pointed'
(10) s became t	siku	tiʔu	'elbow'
(11) mb became m	see (13)	and (15)	
(12) nd became n	tandip	tanipo	'insert'
(13) o following b, mb, d, nd, g, became u	boli kombot	buli ʔomuto	'price' 'k.o. jambu / rose hip'
	dolop	dulopo	'dive'
	bondot	bunoto	'rotten'
	agom	wahumo	'submerge'
(14) a following b became o	batu bobag	botu bubohu	'stone' 'strike'
(15) a following mb, d, nd, g, ŋg became e	bomban dalam tandaʔ bogani nanga	bumeŋo delomo tane buheli lange	'disperse' 'deep' 'birthmark' 'brave' 'nangka fruit or tree'
(16) mp became mb	ompu	wombu	'grandchild'
(17) nt became nd	untu	wundu	'carry on the head'
(18) ŋk became ŋg	paŋkul	paŋgulo	'nit'

5. THE POSITION OF THE BUNDA DIALECT

The Bd dialect is in an intermediate position between G and BM as regards the G sound changes listed above. Several of them have operated in Bd too, but some have not. There is, moreover, one sound change which is peculiar to Bd.

Because of the insufficiency of the quantity and reliability of the data available for Bd this dialect is discussed here briefly and in a very preliminary way. Several uncertainties have remained, some of which may be clarified when better data become available. But some may also be due to the influence of the G language, as G has a dominating position as regards Bd because of its much larger number of speakers and its cultural prestige. It is clear, however, that Bd in several respects is much closer to BM than G. Its intermediate position serves to give additional support to the reconstruction of the G sound changes from a different angle.

1. Final a has not changed, e.g. mata 'eye', but k became ʔ (3) and initial ʔ became w or y (4), e.g. ʔita 'we', yiʔo 'you', wata 'slave'.

2. There are no final consonants because final ʔ was lost (1) and o was added after other consonants (5), e.g. pulu 'ten', wutato 'brother'.
3. The consonants n (6), g (7), and b (8) have not changed, e.g. wana ʔ o 'child', dugi 'thorn', bula 'moon' (see 7.1.1.b. below).
4. But s changed into t (10) and mp into mb (16), nt into nd (17) and ŋk into ŋg (18), e.g. ti ʔ u 'elbow', wombu 'grandchild', bundu - G hundu - BM buntu 'dike', tungudo - G tungudu - BM tu ŋkud 'stick'.
5. Probably mb and nd did not change (11) and (12), e.g. tumbolo - G tumulo 'live', tundu - G tunu - BM tundu ʔ 'show'.
6. Probably o became u following b, d, and g (13) unless these were originally final consonants, e.g. buli 'price', du ŋogo - G du ŋohu - BM do ŋog 'hear', gubi: - G huyi - BM gobii 'night'.
7. Probably a did not change following voiced occlusive (14) and (15), e.g. batu 'stone', dalomo 'deep', bugani 'brave'.
8. r did not change into l (9) but into h, e.g. hibu - G lihu - BM ribu 'thousand', habuto - G lahuto - BM rabut, yabut 'pull out', guhu 'teacher' (Mal guru), tindaho - G tinelo - BM sindar 'shine, light'.
9. Out of the 18 G sound changes 8 operated also in Bd (viz. 1, 3, 4, 5, 10, 16, 17, 18) and 8 did not (viz. 2, 6, 7, 8, 11, 12, 14, 15), one only partially (13) and one was different (9).

6. THE POSITION OF THE BWUOLO LANGUAGE

Because of the paucity of the data available for Bw only a few tentative conclusions can be made concerning the position of this language in comparison with G and BM. They seem to be sufficient, however, to establish that some of the same sound changes have operated in it as in G. For some of these changes only one example was available, however, and for some others none, so that several uncertainties remain to be clarified.

1. Final ʔ was lost (1), e.g. pulu 'ten'.
2. Final a became o (2), and o was added to other final consonants (5), e.g. mato 'eye', ato 'slave', anako 'child'.
3. a following b, d, or g became o, similar to (14) as far as ba is concerned but differing from (15) as to da and ga, e.g. bogu - BM bagu 'new', dolomo 'deep', bugoni 'brave'.
4. o following b, d, or g became u (13), e.g. bugoni 'brave', du ŋogu - BM do ŋog 'hear'.
5. mb became m (11) and nd became n (12), e.g. timo ŋo - BM simba ŋ 'weigh', tonono - BM tonda ŋ 'wages' (cf. G tonelo 'bride price').
6. mp became mb (16) and nt became nd (17), e.g. ombu 'grandchild', gonda ŋo - BM ganta ŋ 'k.o. measuring unit'.
7. As can be seen from several examples mentioned above, k, n, g, and initial glottal stop did not change.
8. There is one change which is peculiar to Bw, viz. that b preceding u became bw (as in the name Bwuolo itself, which is usually known as Buol), e.g. bwuta - BM buta ʔ 'earth'. As a result the difference between *bo and *bu is retained as bu and bwu.

7. SOME IMPLICATIONS AND FURTHER DETAILS OF THE G SOUND CHANGES

In the following paragraphs some further particulars are discussed showing what effects the sound changes have had on Gorontalo.

One of the most important questions is that of their regularity. If these sound changes deserve this name they should be regular or exceptionless, according to the classical definition of the term sound change or sound law in comparative linguistics. It should be determined how regular they are and, if there are or seem to be exceptions, in how far these can be explained. A partial answer to this question can be found in the way they have been formulated. Whereas several of them have been defined as operating unconditionally, others are stated to apply under specific conditions only, viz. initially, finally, and preceding or following particular sounds. The latter may be considered to be exceptions to unconditional sound changes, regularly explained within the context of the sound changes themselves. Some more complicated cases follow below.

In the present section exceptions occurring in loanwords are left aside provisionally, as they show specific differences, some of which class them together as a distinct group. They are discussed in a separate section (see 8, below).

7.1. The word end

7.1.1. If the sound change under (2) operated regularly the result would be that a never occurs word finally. There are nonetheless several words which have a final a. From those cases in which cognates exist in BM three reasons can be given for the occurrence of final a.

a. Some of these words have originally had a final ʔ , which was lost according to (1), so that only after this loss of final ʔ the originally prefinal a became final, e.g.

dila – BM dila ʔ 'tongue' (1)
 dupa – BM dupa ʔ 'to forge' (1)
 huta – BM buta ʔ 'earth' (8), (1)
 $\text{ʔ}o/yinda$ – BM inta ʔ 'one' (4), (17), (1)
 mama – BM mama ʔ 'chew betel' (1)

b. In other cases, many of which are tri- or quadrisyllabic, a final n was lost after prefinal a. Although there are several examples, this loss of final n can not be called a sound change in the same right as the others because there are other words which end in -alo as a reflex of -an, such as tulalo – BM tulan 'bone'. Possibly this loss of final n was in fact the (sporadic) loss of final lo after a, e.g.

dutula – BM dutunan 'river' - 'river for bathing' (6)
 hulawa – BM bulawan 'gold' (8)
 hutiya – BM gosi $\text{ʔ}an$ 'rotan' (7), (13), (10)
 $\text{ʔ}olowala$ – BM koloanan 'right hand' (3), (6)
 $\text{ʔ}ologiya$ – BM koronjan 'nobleman' (3), (9)
 tambala – BM tampalan 'spotted' (17)
 wulipa – BM ulipan 'centipede' (4)

In at least one case there is a semantic differentiation between the words with and without the final n or lo, viz. hulalo 'moon, month' and hula 'month (as a measuring unit)', e.g. wolomo hula 'six months' - BM onom no bulan.

In Bd this loss of final n after a seems to be more regular or perhaps even exceptionless, e.g. tula 'bone', bula 'moon', dala 'road'.

A similar loss of final lo from n is met with in the verbal suffix -a (from -alo, from -an), see sub 10.7.3. below, and may be regarded as the origin of the loss treated in the present subsection.

c. Occasionally we find other a final words which have lost a final consonant other than l from n, e.g. hula - BM bulag 'albino'.

7.1.2. Originally final a[?] did not result in final a when this was preceded by mb, d, nd, or g because a changed into e in this position according to (15). There are only examples of words with originally final nda[?], ga[?], and probably mba[?], viz.

tane - BM tanda[?] 'birthmark' (12), (15), (1)
 bohe - BM бага[?] 'abscess' (14), (7), (15), (1)
 tohe - BM toga[?] 'Lamp' (7), (15), (1)
 time - (e.g. Sangirese) timba[?] 'scoop out' (11), (15), (1)

7.1.3. The o which according to (5) was added after final consonant became u according to (13) when this consonant was b, d, or g. This also happened when final a had changed into o according to (2), while the preceding consonant was b, mb, (n)d, or g. As a result there are no words ending in *bo, *do, *ho (from go), *mo (from mbo), or *no (from ndo). Instead these words end in bu, du, hu, mu, or nu, e.g.

wangubu - BM aṅkub 'cover' (4), (18), (5), (13)
 wulodu - BM ulod 'caterpillar' (4), (5), (13)
 yilihu - BM ilig 'stream' (4), (7), (5), (13)
 pobu - BM poba 'burn' - 'burn lime' (2), (13)
 bohu - BM бага 'coal' (14), (7), (2), (13)
 to/onu - BM onda 'where' (12), (2), (13) (to 'in, at')
 wonu - BM onda 'if' - 'where'

No examples have been found of -du from -da, and of -mu from -mba.

7.2. Initial glottal stop and initial semivowels

Information from native speakers leaves no doubt that glottal stop occurs in initial position in a number of G words. According to sound change (3) it replaces k in that position as it does in intervocalic position. As a result of (3) and (4) vowels do not occur word initially in G.

It is less clear whether initial glottal stop also occurs in BM. According to Dunnebier (1929:317) an initial vowel is preceded by a 'soft blow' (*lichte aanblazing*) to be distinguished from glottal stop as a 'real consonant'. But M. Charles (1974:486) transcribes such BM words with initial q for glottal stop

presumably on the basis of information from BM speakers. If nevertheless the former alternative is the correct one, the G sound change (4) might preferably be formulated as: y was added before initial i and w before other initial vowels.

It seems clear, however, that this sound change took place in word-initial position only, as can be seen from examples as BM *onda* 'where' - G *wonu* 'if' and G *to:nu* 'where', in which the final o of the preposition *to* and the initial o of **onu* have become o:; and BM *ompu* 'grandchild, grandfather' - G *wombu* 'grandchild' and *tiyombu* 'grandfather', in which the semivowel is not w but y, following the final i of the 'personal article' *ti* (from *si*, as in Malay, but not in BM). The same is the case in G *ti:lo* 'mother' and *tiyamo* 'father' from **ina* and **ama* (BM *ina?* and *ama?*, contra 7.1.1.a.). Similarly the term *wali-wali* 'k.o. social class' may be pronounced with two w's but also as *waliyali*.

7.3. Convergence and homonymy

In several cases two or more sound changes have directly or indirectly led to the same result, though starting from different origins. In consequence there is a marked tendency towards convergence of sounds in G, which in some cases resulted in the emergence of new homonyms. Some examples are given below.

Since both g and b (before u) changed into h, initial hu in present-day G may be the reflex of either gu or bu, e.g.

huwoto - BM *gu?ot* 'gums' (7), (4), (5)

huwo?o - BM *buok* 'hair' (8), (3), (5)

Final hu may be the reflex of ga, gu, go, or bu, e.g. (with two pairs of new homonyms):

bohu - BM *baga* 'coal' (14), (7), (2), (13)

bohu - BM *bagu* 'new' (14), (7)

wahu - BM *agow* 'to rob' (4), (7), (13) (BM -ow from -o)

wahu - BM *abu* 'ashes' (4), (8)

Similarly both n and r changed into l, so that l in G may be the reflex of n, r, or l, e.g. (with two sets of new homonyms):

huli - BM *buli* 'loose' (8)

huli - BM *buni* 'conceal' (8), (6)

huli - BM *buli?* 'underside' - 'buttocks' (8), (1)

huli - Bd *guli* 'to install' (7)

mulo - BM *muna* 'first' (6), (2)

mulo - BM *mula* 'to plant' (2)

tuludu - BM *tulud* 'push off' (5), (13)

tulubu - BM *turub* 'burn' (9), (5), (13)

Other sound changes may also bring about new homonyms, e.g.

tiyo - BM *sia* 'he' (10), (2)

tiyo - BM *siow* 'nine' (10)

7.4. Some place names

7.4.1. The G pronunciation of names of places outside the G territory naturally shows the effect of the G sound changes and presents good evidence if the original names are known. Some examples are the following:

Moŋonu - BM Moŋondow (12), (13)

Moladu - BM Monado (Manado or Menado, the capital of Minahasa) (6), (13)

Huwa - Gowa (Makassarese kingdom in South Celebes) (7), (13)

Huwolo - Bw Bwuolo - BM Buol (8), (5)

?Uwanenŋo - Kuwandaŋ (place on the north coast) (3), (12), (15), (5)

Bodu - BM Bado? (the Bajo sea-nomads) (14), (13), (1)

Bune - Bd Bunda (12), (15)

The last comparison shows that the original name must have been *Bonda?, in which the changes (13) and (1) have operated both in G and in Bd. If Bunda was the original form, the G one should have been *hunu.

7.4.2. The G name for Gorontalo is Hulondalo, which can be explained as the regular result of the sound changes (7), (13), (9) and (17). This is partly but not completely confirmed by the Bd and BM forms of this name. The origin of the first l is uncertain. It can be the G reflex of l, r, or n, which in Bd are reflected as l, h, and n. The Bd form of the name is Golondalo, however, and the BM form is Gonotalon. This conflicting evidence can be explained by assuming that the BM form is the original one and the Bd one is a borrowing from G dating from a time when n had become l and nt had become nd in G but the other G sound changes concerned had not yet taken place. The final n of the BM form, which may have dropped in both G and Bd, seems to confirm that the BM form was the original one. The possibility that at one time the name was pronounced Golontalo is confirmed by its form in Makassarese, which is Golontalo.

7.5. Semantic change

Although the G - BM cognates given as examples have been limited to those which show no or only a small semantic difference, there are several others in which semantic change has clearly taken place, e.g.

bilohu 'to see' - BM bilog 'blind' (7), (13)

yilaluhu 'town' - BM inalug 'village road' (4), (6), (7), (5), (13)

?ulu?u 'hand' - BM konuku 'nail of the finger' (3), (6) (with assimilation of o into u)

walito 'animal skin' - BM alit 'rope made of animal skin' (4), (5)

8. BORROWINGS

There are quite a few Gorontalo words which were borrowed from other languages such as Arabic, Malay/Indonesian or Dutch. Very few, if any, of these borrowings show the effects of the G sound changes whenever these are applicable. Most of them do so only partially or not at all. In many cases one particular sound change is applied but another one not. Some sound changes hardly ever occur in loanwords, others occur in some words but not in others. Sometimes there

are two variants of a loanword, one with and one without the sound change concerned. There is also a difference depending on the educational level of the speaker, in such a way that the knowledge of Malay/Indonesian tends to increase the number of cases in which the sound changes are not applied in borrowings from this language.

8.1. Loanwords which have been affected by every applicable sound change are scarce. Even the following two examples pose some problems.

tihi 'mosque' - BM sigi (10), (7), from Malay masigit 'mosque' which is ultimately from Arabic masjid 'place of prostration'. There is a semantic difference in that in BM sigi is a (non-Muslim) 'small village temple' and masigi is 'mosque'. Nevertheless it seems most plausible that both tihi and sigi are loanwords. wadala 'horse' (not in BM; cf. BM kabalo 'horse' from Spanish) is ultimately derived from Jav. *ajaran 'trained animal' (4), (9), d from j, and final n dropped according to 7.1.1.b. above.¹ The a instead of e after d is an exception to sound change (16), possibly caused by the palatal origin of this d.

8.2. The way in which various G sound changes have affected borrowings is further exemplified below.

8.2.1. Final a hardly ever changed into o according to (2). An example is perhaps tolimo 'receive', if this is a loan from Malay tarima (cf. BM tarima explained as borrowed from Malay) and not inherited.

8.2.2. Initial ʔ is hardly ever replaced by semivowel according to (4); wadala 'horse' is the only example I have found. As a result there are several loanwords with initial vowel (preceded by ʔ) which is not the result of the loss of initial k, e.g. ʔalabaʔa 'Wednesday' (Mal/Ar arbaʔa) (Pateda: arabaʔa). On the other hand there are some examples of the loss of initial k according to (3), e.g. ʔapitalau 'admiral' (Mal kapitan laut), ʔatetela 'ketela', ʔupiya 'kopiah, fez'. In many words the k has been preserved, e.g. kaputenji 'captain' (Du kapitein), nika 'marry' (Mal/Ar nikah).

8.2.3. Hardly any borrowing has o added after final consonant according to (5). An example is perhaps kayito 'hook' (Mal kait), which exists beside ʔayito 'hook' (cf. BM kait 'to knit').

In most other cases the vowel added after final consonant of loanwords is either i (unless the final consonant is a labial) or u, while at the same time the penultimate vowel is usually lengthened. Examples with final i are: basi 'craftsman' (Du baas), di:kili (Mal/Ar dikir), ka:pali 'ship' (Mal kapal), ka:yini 'cloth' (Mal kain), ku:raŋi 'less' (Mal kuraŋ), ka:ntori 'office' (Mal/Du kantor), me:maŋi 'certainly' (Mal memaŋ), pa:tali 'market' (Mal pasar),

ra:cuni 'poison' (Mal racun), sabari 'patient' (Mal sabar). The vowel is u if the final consonant is a labial, e.g. ?ima:mu 'imam', capu 'seal' (Mal cap), sababu 'because' (Mal/Ar sabab), lemu 'glue' (Mal/Du lem), bomu 'harbour' (Du boom), kolopu 'all right' (Du klopt), but also in some other cases, e.g. boku 'bay' (Du bocht), roku 'frock' (Du rok), bo:roku 'bail' (Du borg), ba:ngurutu 'bankrupt' (Du bankroet), bu:tulu 'bottle' (Mal botol).

8.2.4. There are few examples of n which changed into l according to (6), e.g. talala 'trousers' (Mal celana).

8.2.5. Often but not always g changed into h according to (7), e.g. da:hanji 'commerce' (Mal dagan), but da:giŋi 'meat' (Mal daging), halati 'glass' (Mal/Du gelas), hula 'sugar' (Mal/Skt gula), huna 'utility' (Mal guna), naha or naga 'dragon' (Mal/Skt naga), lahu 'song' (Mal lagu), but ga:risi 'line' (Mal garis), etc.

8.2.6. Mostly r changed into l according to (9), e.g. gulu 'schoolmaster' (Mal/Skt guru), halahadi 'saw' (Mal gergaji), luhi 'loss' (Mal rugi), su:kali 'difficult' (Mal sukar), but also guru, rugi occur.

8.2.7. There are some examples of the change of bu into hu according to (8), e.g. sahatu 'Saturday' from *sabutu (Mal/Ar sabtu) and tahuda 'word' from *sabuda (Mal/Skt sabda).

8.2.8. Usually s changed into t according to (10), e.g. ka:tulu 'mattress' (Mal kasur), tababu 'because' (Mal/Ar sabab), tu:kali 'difficult' (Mal sukar), though in many cases s remained.

8.2.9. There are few examples of the change of o into u following b, d, or g according to (13), e.g. bu:tulu 'bottle' (Mal botol), du:bulu 'double' (Mal/Du dobol), dusa 'sin' (Mal/Skt dosa), huhuhu 'grand vizier' (BM gogugu from Ternatan (jou) gugu).

8.2.10. In most cases mp changed into mb, and ŋk into ŋg according to (16) and (18), e.g. pomba 'pump' (Mal/Du pompa), ?o:ŋgosi 'expenses' (Mal/Du oŋkos), pa:ŋgati 'rank' (Mal paŋkat).

8.3. The loanwords in G as briefly described above constitute a heterogeneous group of words, as is usually the case in any language. As they were adopted from structurally quite different languages and presumably at different times in the past, they have been adapted to the existing G sound structure in various ways and degrees. This in itself is a general characteristic which classes them together as a separate group. A second more specific one is the fact that, as mentioned sub 8.2.3. above, the vowel added after final consonant is not o according to the earlier sound change (5) but i or u according to more or less strictly applied rules.

At the same time many examples of loanwords in which G sound changes did take place, show most clearly that the particulars of these sound changes were the same as those found in the inherited part of the vocabulary.

9. UNEXPLAINED EXCEPTIONS

There are a number of G words which, while not obviously recognisable as borrowings, do not show the effect of the applicable sound change as it should have been expected from comparable cases. Most conspicuous are those which clearly have cognates in BM. They do not invalidate the evidence for the sound changes as it has been given earlier, but should be regarded as unexplained exceptions as long as no acceptable explanation has been found.

We give some examples of the most obvious cases.

9.1. Some words have initial ha-, which should have become he if it reflected ga according to (15) as in e.g. hetuto - BM gatut '*hundred*', viz.

hama - BM gama? '*take*'
hawu - BM gau '*cigarette*'

9.2. A number of G words have initial da, which should have become de according to (15) as e.g. in deheto - BM dagat '*sea*', viz.

dalalo - BM dalan '*road*'
damahu - BM damag '*resin*'
dalahu - BM darag '*yellow leaf*' - '*yellow*' (but G lalahu Bd dahago '*yellow*') cf. also datahu '*flat, level*', not in BM but clearly from *datag.

9.3. A number of G words have initial ba, which should have become bo according to (14) as e.g. in botulo - BM batun '*ascend*'. Some of them have cognates in BM, viz.

bataŋa - BM bataan '*body*' - '*personal appearance*'
bataŋo - BM batak '*stem*'
bata'o - BM batak '*bad-mannered*'

9.4. A number of G words have one or more k's for which ? should have been expected according to (3). Some of them have BM cognates, viz.

kikilo?o - BM kilok '*tickle*'

kokobu - BM kokob '*bite*' - '*chew something hard*', cf. e.g. ?o?odu - BM kokod '*embrace*'.

10. AFFIXES AND MORPHOPHONEMIC ALTERNATIONS

10.1. Quite a few of the G verbal and nominal affixes show the effect of one or more of the sound changes, leading in some cases to morphophonemic alternations. We give the following examples.

10.1.1. The verbal prefix mo has several allomorphs, among which in the first place moN- may be mentioned. The nasal replaces the initial consonant of the base to which the prefix is added and is either m or ŋ. The m occurs if the initial consonant of the base is p, b, or h; the ŋ only if the initial consonant is k, ?, w or y, e.g.

pututo : momututo '*to wrap up*' (BM putut)

biyahu : momiyahu '*to care for*' (BM biag : momiag)

huwato : momuwato '*to lift*' (BM buat : momuat '*to lift*', '*to go upwards*')

kokobu : moŋokobu '*to bite*' (BM kokob : moŋokob '*to chew something hard*')

?u?udu : moŋu?udu '*to remove scales*' (BM kukud : moŋukud '*to scrape*')

wuli : moŋuli '*to revert to previous condition*' (BM uli? : moŋuli '*to recover*')

yilu : moŋilu '*to drink*' (BM inum : moŋinum).

The cases in which initial h is replaced by m are restricted to those in which h is followed by u and are therefore explicable, as in the example given, by the sound change (8) (bu became hu). In the same way the replacement of initial ?, w and y by ŋ can be explained by the sound changes (3) (k became ?) and (4) (? became w or y), respectively.

10.1.2. Most bases with initial w and y, however, have an allomorph of the prefix, in which the initial consonant is replaced by h, while a following initial w is replaced by e, and o following initial w by u. These replacements are explicable by the sound changes (7) (g became h), (15) (ga became ge), and (13) (go became gu), as is confirmed by BM cognates, e.g.

yimo?o : mohimo?o '*to collect*' (BM imok : mogimok)

wulilo : mohulilo '*to steer*' (BM ulin : mogulin)

waluto : moheluto '*to invite*' (BM anut : moganut)

wotuto : mohututo '*to pass gas*' (BM otut : mogotut)

This implies that diachronically the h did not replace initial w or y, since the latter never occurred in this position, which in the prefixed verb was not the initial one. In fact the allomorph moh- is a reflex of the prefix mog-, which occurs in BM, where it is prefixed to vowel-initial bases only. Clearly the same limitation of occurrence once applied to moh- in G and therefore may be ascribed to Proto-G-BM.

Both G moh- and BM mog- alternate with moŋ- in such a way that every given base concerned has unpredictably either the former or the latter prefix (in BM sometimes both occur with the same base with semantic difference). Sometimes a difference can be observed in this respect between G and BM, e.g. wunemo : mohunemo 'to cure' - BM undam : moŋundam.

10.1.3. Bases with initial t have an allomorph of the prefix in which t is replaced by l. This can be explained by the assumption that originally this t was replaced by its homorganic nasal n, which afterwards changed into l according to (6), e.g. toneʔo : moloneʔo 'to take soundings' - BM tondak : monondak.

10.2. Similarly the preterital prefix lo- reflects no- (BM idem) as a result of sound change (6).

10.3. The prefix ʔo- denoting possession, which is not dealt with in Badudu 1982 and treated as an independent particle in Pateda 1977, is the reflex of ko- according to (3), e.g. pali 'wound' : ʔopali 'wounded' - BM paliʔ : kopaliʔ.

Formally the same prefix occurs with numerals, e.g. wopato 'four' : ʔo:pato 'fourth' - BM opat : koopat.

10.4. The prefix moʔo-, meaning 'to be able to', is the reflex of moko- (BM idem) according to the sound change (3), e.g. moʔoduŋohu 'to be able to hear' - BM mokodoŋog.

10.5. The prefix mohi-, meaning 'to wear (a piece of clothing)', is the reflex of mogi- (BM idem) according to sound change (7), e.g. mohisapatu 'to have shoes on' - BM mogisapatu.

10.6. There are two infixes, the passive/preterital -il- and the futural -um-. The former is the reflex of -in- according to (6). It may be infixes within prefixes, e.g. ʔo- : ʔilo- (BM ko- : kino-). When it is infixes in a base which has initial w, this consonant is replaced by y according to sound change (4), e.g. wohi : yilohi 'give' (BM ogoi : inogoi).

In those cases in which the change of an initial consonant is conditioned by the occurrence of a particular vowel immediately following it, or the change of vowel by the occurrence of a specific consonant immediately preceding it, such as according to (8) (bu became hu), or (13) (do became du), it might have been expected that these changes did not occur if an infix was placed between this consonant and the immediately following vowel. Some examples of this expected phenomenon occur.

When we compare G dumodupo 'morning' and du:dupa 'early in the morning' with BM mododop 'early', it is clear that do changed to du twice according to (13), but not where d and o were separated by the infix -um-. Similarly, when

comparing G bubohu 'hit' and bilobohu 'hit (pass.)' with BM bobag 'hit', we see that bo changed to bu according to (13), but not where b and o were separated by the infix -il-. Other examples are: buli 'price' but biloli 'debt' - BM boli 'debt, price', and hutu 'penis' but bilutuwa 'castrated' - BM butu? 'penis'. G hulo'o 'to sit' has no cognate in BM nor in Bd, and the initial h could therefore have originated from b or from g. That the former was actually the case appears from the noun bilulo?a 'seat' (not in Pateda, but found in a G poem).

Probably these are exceptional occurrences, however, at least in the present-day language. Usually this kind of sound change also applies when an infix separates the consonant and the vowel concerned, e.g. dehito 'snatch' : dumehito 'will snatch' - BM dagit : dumagit.

10.7. The suffix -a occurs in various morphological functions, among which the imperative may be specially mentioned here. As to its morphophonemic aspects, however, it can be treated as one and the same suffix in all its functions.

10.7.1. It shows the effect of sound changes in cases of bases to which o was added after final consonant according to (5). When the suffix was present, the final consonant of the base was no longer final and therefore o was not added here, or, in other words, the suffix was added immediately after the final consonant of the base. The synchronic result is that the final o of such bases is replaced by the suffix, e.g. pututo 'wrap up' (BM putut) : pututa 'wrap up!'.

This also happened when the final consonant of the base was b, d, or g. The suffix -a was added immediately after these, but then changed into e according to (15), just as the o added after the final consonant of such bases changed into u according to (13). The synchronic result is that the suffix -a appears as -e replacing final u, e.g.

tulubu 'burn' (BM turub) : tulube 'burn!'
 tuludu 'push' (BM tulud) : tulude 'push!'
 tunuhu 'follow' (BM tundug) : tunuhe 'follow!'

10.7.2. In some cases the suffix is synchronically not -a but -ma, -la, or even -ala. These exceptional cases can be explained as the result of the loss of the final consonant of the base (together with that of the o added after it and, in the last mentioned case, of the a preceding it) in other derivations. There are at least three examples which have cognates in BM showing that this was what happened, viz.

yilu 'drink' : yiluma 'drink!' - BM inum
 wawo 'weave' : wawola 'weave!' - BM abol (with assimilation of b to w ?)
 pota 'carry on the shoulder' : pota:la 'carry on the shoulder!' - BM potaan.

10.7.3. The fact that this -a suffix did not change into -o according to (2), although it is a final vowel, must mean that it was not yet final when this sound change took place and that a final consonant following it was lost

afterwards. This final consonant must have been n and the suffix in fact was the well-known suffix -an, as is proved by BM, where -an occurs in many cases with the same morphological functions as -a and its allomorphs in G, among others that of the imperative. Some examples out of many are the following:

yiluma 'drink!' - BM inuman 'drink of it!'
 litode 'twine!' - BM litodan 'roll up!'
 pitota 'tighten!' - BM pitotan 'strangle (with a rope)!'

This loss of the final n is analogous to that in trisyllabic nouns such as hulawa 'gold' (BM bulawan) mentioned earlier (see sub 7.1.1.b. above).

As regards the -a suffix as imperative it may be concluded that this, contrary to Adriani's statement (1908:163) on this point in his remarks about the Gorontalo language, is not the 'conjunctive' or 'irrealis' suffix -a, which is found in several languages such as Javanese, Makassarese and some Bisayan dialects with imperative meaning among others. If it existed in G, it should have become -o according to the sound change (2).

The reason why the suffix -an lost its final n in G may have been that it regularly should have become - and possibly first did become - -alo, whereas -lo also exists as a suffix as well as a separate particle. It may occur in combination with the imperative -a with emphatic meaning, e.g. tali 'buy' : talia 'buy!' : talialo 'do buy!'.

10.8. The 3rd person possessive suffix has two allomorphs, viz. -iyo and -liyo, which reflect -*ia and -*nia according to the sound changes (2) and (6). In BM they are -ea and -ña, the former occurring after final consonant (except ?), the latter after final vowel (and ?). The same distribution (? left aside) must have existed in G prior to the addition of o after final consonant according to (5). As a result the allomorph -iyo is suffixed to words ending in this o but then replacing this final vowel, e.g. wala'o 'child' : wala'iyc 'his child', but e.g. wato 'slave' : watoliyo 'his slave'. Historically the o of sound change (5) was never added to words like wala'iyo since there was no final consonant because of the presence of the suffix.

This phenomenon can be used as a means to find out whether or not a final o or u is diachronically an added one according to (5). For example, the assumption that bi:hu 'lip' is the cognate of BM bibig according to (5), (7) and with loss of the second b, is confirmed by bi:hiyo 'his lip'.

11. RELATIVE ORDER

If the several sound changes are compared among each other, it can be remarked that some of them are interrelated with certain others in that they concern the same phoneme in different ways. If these interrelated sound changes are analysed under the assumption that sound changes operate exceptionless within their respective conditional limits and within a certain limited time, it appears to be possible to establish the relative order in which such interrelated sound changes took place.

11.1. In some cases a certain phoneme was lost in all its occurrences as the result of one sound change, but the same phoneme re-emerged elsewhere as the result of another sound change, e.g. all n's were lost after they had become l according to (6), but n re-emerged when d in nd had become lost according to (12); similarly initial ʔ was lost after it had become w or y according to (4) but re-emerged in other words as the result of the change of k into ʔ according to (3). This implies that these changes occurred in both cases in the order in which they have been mentioned just now, i.e. the change of n into l took place and ceased operating before that of nd into n began operating, since otherwise n from nd would also have become l. Similarly the change of ʔ into w/y predated that of k into ʔ.

11.2. In some cases it can be observed that more than one sound change must have taken place in one and the same position in a given word, which implies that one of these changes operated later in time than the other. Their relative order follows from the details of the changes concerned. For example, if bohu is the reflex of *baga, the final a can only have become u, if it changed into o according to (2) first, since otherwise a following g would become e according to (15). This means that the change of final a into o (2) predated that of go into gu (13) and that of ga into ge (15).

Another example is onu in to:nu 'where', BM onda. The final a became o according to (2) first, then changed into u following nd according to (13), after which nd became n according to (12), the relative order thus being onda > ondo > ondu > onu.

A comparable case is the following. If tohe 'lamp' is the reflex of toga⁷ (BM idem), the loss of final ʔ according to (1) must have postdated the change of final a into o according to (2), since otherwise the a in ga⁷ would have become final and would then have changed into o, and this into u following g according to (13). This means that the change of final a into o (2) predated the loss of final ʔ as well as the change of go into gu (13).

11.3. There are also sound changes which must have taken place at the same time, such as (13), o following (m)b, (n)d, or g became u, since this can be formulated as: o following voiced stop became u. This implies that the change of g into h (7) must have postdated (13), since otherwise the final syllable of *bago (from *baga) would have become *ho instead of hu. For the same reason the change of ba into bo (14) must have postdated (13), since otherwise the first syllable of bohu (from *baga) would not have remained bo but would have become bu.

Similarly the change of bu into hu (8) must have predated (13), since otherwise bu from bo (13) would also have changed into hu.

Similarly the change of a following mb, (n)d, or g into e (15) must have predated that of g into h (7), since otherwise *toga would have become *toha instead of tohe.

But the change of ba into bo (14) must have predated (15), since otherwise ba would also have become be according to (15), and it must have postdated (13), since otherwise bo from ba would have become bu according to (13).

11.4. The changes of mb to m and nd to n (11) and (12) are interrelated as in both cases voiced stops preceded by homorganic nasal were lost, but apparently there was no analogous change of ŋg into ŋ (cf. *lange* from **naŋga* according to (15). Examples such as *kombot* - *ʔomuto* '*k.o. jambu / rose hip*', *tondok* - *tonuʔo* '*fence*', as well as *bombarŋ* - *bumeŋo* '*disperse*' and *sindar* - *tinelo* '*light*' show that these changes postdated both (13) and (15).

The changes of mp into mb, nt into nd, and ŋk into ŋg (16), (17) and (18) clearly postdated those of mb into m and nd into n (11) and (12), since otherwise the voiced stops resulting from the former would have become lost as a result of the latter.

11.5. It follows from the foregoing subsections that there are three pairs of interrelated sound changes as well as a group of seven, all of the latter involving voiced stops, for which the relative order within the pair or group is certain, namely:

- A. 1. final a > o (2)
- 2. loss of final ʔ (1)
- B. 1. initial ʔ > w/y
- 2. k > ʔ (3)
- C. 1. n > l (6)
- 2. nd > n (12)
- D. 1. bu > hu (8)
- 2. (m)bo > (m)bu, (n)do > (n)du, go > gu (13)
- 3. ba > bo (14)
- 4. mba > mbe, (n)da > (n)de, ga > ge (15)
- 5. g > h (7)
- 6. mb > m, nd > n (11), (12)
- 7. mp > mb, nt > nd, ŋk > ŋg (16), (17), (18)

The relative order of A, B, C, and D cannot be ascertained more precisely than that A1 predated D2, but D4 (and thus D1-3) predated A2 (see 7.1.2.), whereas C1 predated D6.

12. SUBGROUPING

Adriani (1914:183-217) and, following him, Esser (1938) have placed Bolaang Mongondow in a subgroup of Philippine languages together with Sangirese-Talau and the Minahasa languages, and have grouped Gorontalo together with Buol, Kaidipan and Bulango in a Gorontalo subgroup not belonging to the Philippine languages.

In view of the materials contained in the present article it seems warranted to revise this classification and to group G and BM much more closely together. If BM should be recognised as belonging to the Philippine family of languages, G should at least be included into it too.

This conclusion has also been reached by M. Charles who included Gorontalo in his treatment of Philippine languages "because comparison revealed striking similarity in the vocabularies and (to a point) in the phonological histories of Mongondow and Gorontalo" (1974:487). And we may

add similarity in their morphologies as a third indication of their close relationship.

Whether G and BM together with the other languages mentioned by Adriani and Esser belong to a common subgroup, must be determined by further comparison of these languages among one another and with what is known of Proto-Philippine and Proto-Austronesian.

We mention some points concerning G and BM as examples.

12.1. There are other G-BM correspondences not mentioned before. For instance, in several examples G final *-e* corresponds to BM final *-oi*, e.g.

bele - BM baloi 'house' (with assimilation of the first vowel to the second, cf. Bw bole 'house')
 pate - BM patoi 'kill'
 tuŋe - BM tuŋoi 'horn'
 wale - BM anoi 'ant' (4), (6)
 wu:wate - BM uatoi 'iron'
 pale - BM payoi 'rice plant'

In some cases, however, the correspondence is G *-i* : BM *-oi*, e.g.

pani - BM pandoi 'skilled' (12)
 wohi - BM (b)ogoi 'give' (4), (7)

12.2. Though in many cases BM did not change from putative Proto-G-BM, there is also evidence for sound changes which have taken place in BM. For instance, Proto-G-BM apparently retained PAN *S- as h-, with subsequent loss of h- in BM, but not in G, e.g.

G hataŋo - BM asaŋ 'gills'
 G huwabu - BM uab 'yawn'

In several cases of the correspondence of G *l* : BM *r* there is a BM doublet with *y*, which may point to a development in BM of *r* to *y*, e.g.

G lahuto - BM rabut / yabut 'pull out' PAN (r)ebut/Rabut
 G le:to - BM raat / yaat 'bad' PAN zaqet
 G hulu?o - BM buruk / buyuk 'rotten' (cf. Bd buhu?o) PAN buRuk
 G lolom/bula - BM bura? / buya? 'foam' PAN bujaq

In several other cases, among which the reflexes of PAN *j*, the correspondence is G *l* : BM *y* : Bd *h*, e.g.

G tili - BM si(y)i? - Bd tihi 'side' PPH siDiq
 G tolomo - BM toyom - Bd toho 'ant'

12.3. As an alternative of the G sound change (13), i.e. *o* following (m)b, (n)d, *g* became *u*, it may be assumed that this *u* is a direct reflex of an inherited or secondary schwa, which in other environments is reflected as *o*, whereas it is reflected as *o* in all environments in BM. This assumption is made by Charles (1974:469-470) in his explanation of the G word in the example "PPh *dapdap

'Erythrina spp.' ... BM dodap ...; G dudepo ... with /u/ < secondary *e following *d". These alternatives illustrate the difference between considering G and BM as reflecting the higher level proto-languages PPh or PAN directly or via Proto-G-BM, as has been done in the present paper. It seems, however, that examples such as BM poba - G pobu show that the G sound change (13) has indeed taken place, since the final o of the putative intermediate form *pobo does not reflect a schwa. Examples of the change of o into u following g are also to be found in such loanwords as the toponym Huwa from Makassarese Gowa and huhuhu 'grand vizier' from gogugu (BM idem), from Ternatan (jou) gugu.

NOTE

1. In Jav. itself *ajaran 'horse' from ajar 'teach, train' has changed into jaran with loss of initial a analogically to Jav. loss of the verbal prefix a-. The same change may have taken place in Makassarese, in which 'horse' is jaraŋ and where the verbal prefix a- was usually also dropped. In Buginese, where this prefix has not dropped, the word for 'horse' has kept its initial a, whereas the palatal consonant was nasalised and geminated: aññaraŋ.

BIBLIOGRAPHY

ADRIANI, N.

- 1908 Breukink's bijdragen tot eene Gorontalo'sche spraakkunst besproken. *Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 70:150-165.

ADRIANI, N. and A.C. KRUYT

- 1914 *De Bare'e-sprekende Toradja's van Midden-Celebes*, vol.3. Batavia: Landsdrukkerij.

ANDEL, W.J.D. van

- 1929 Boeoolsche adatrechttermen. *Adatrechtbundels* 31:28-68.

BADUDU, J.S.

- 1982 *Morfologi kata kerja bahasa Gorontalo*. (Doctoral thesis in linguistics, University of Indonesia, 1975.) Jakarta: Panerbit Djambatan.

CHARLES, M.

- 1974 Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics* 13:457-509.

DUNNEBIER, W.

- 1929-30 *Spraakkunst van het Bolaang Mongondowsch. Bijdragen tot de Taal-, Land- en Volkenkunde van Nederlandsch-Indië* 85:297-468, 524-621; 86:42-177. Published, same years, in book form, The Hague: M. Nijhoff.
- 1951 *Bolaang Mongondowsch-Nederlandsch woordenboek met Nederlandsch-Bolaang Mongondowsch register*. The Hague: M. Nijhoff.

ESSER, S.J.

- 1938 Talen (map 9b). In: *Atlas van tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

KASIM, M.M., Habu WAHIDJI, Mansoer PATEDA, Husain JUNUS, and Nani TULOLI

- 1979 *Struktur bahasa Suwawa*. Gorontalo: Laporan Penelitian.

PATEDA, Mansoer

- 1977 *Kamus bahasa Gorontalo-Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- 1981 *Kaidah bahasa Gorontalo*. Gorontalo: Viladan.

SCHRÖDER, E.E.W.G.

- 1908 *Gorontalosche woordenlijst*. The Hague: M. Nijhoff.

THE INADEQUACY OF THE INERTIAL DEVELOPMENT PRINCIPLE IN ACCOUNTING FOR SOUND CHANGES IN SEVERAL AUSTRONESIAN LANGUAGES

FARID M. ONN and Mangantar SIMANJUNTAK

1.0. James Foley's inertial development principle (henceforth, IDP) in phonology (Foley 1970a,b, 1973, 1977) has been applied to account for sound changes in certain European languages (mostly Germanic) by Foley himself, in Scots by Taylor (1973), and in several Austronesian languages (Malay, Indonesian, Javanese, Toba Batak, Tagalog, and Samoan) by Mashudi Kader (1979). While in certain Germanic languages the IDP seems to be, at first glance, reasonably adequate in explaining some aspects of sound change, in Austronesian languages, however, this principle is totally inadequate. Thus, the universal validity of the IDP that has been claimed by Foley and subscribed, among others, by Mashudi should be seriously questioned, especially with regard to sound changes in the above Austronesian languages.

In this paper we seek to show that the IDP is in many cases not only inadequate but somewhat strange in handling sound changes in the data offered by Foley and Mashudi, and even Taylor (1973). We would also like to argue that sociolinguistic as well as psycholinguistic aspects should not be excluded from any attempt at solving the problems of sound changes in any language (cf. Weinreich, Labov, and Herzog 1968; Slobin 1977; Baron 1977; Andersen 1973; Goyvaerts 1975; Simanjuntak and Ramii Salleh 1980), including the Austronesian languages. Further we will argue that the principles used to account for sound changes in one particular language may not be valid for other languages due to, among other things, different sociolinguistic aspects. However, before we proceed to discuss the application of the IDP to the Austronesian languages to see its adequacy/inadequacy, let us briefly sketch out what the IDP is all about.

2.0. Roughly the IDP (Foley 1977:107) states that:

- (1) strong elements strengthen first and most extensively and preferentially in strong environments, and
- (2) weak elements weaken first and most extensively and preferentially in weak environments.

The α and β scales of relative phonological strength of segments of the IDP

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 263-284. *Pacific Linguistics*, C-75, 1982

© FARID M. ONN and Mangantar SIMANJUNTAK

is as follows (Foley 1977:34):

(1)	4	↑	kk	tt	pp
	3		k	t	p
β	2		g	d	b
	1		γ	ð	β
			1	2	3
			α		

The β strength scale arranges consonants according to manner of articulation and voicelessness/voicedness, whereas the α strength scale arranges consonants according to their place of articulation. Thus, following the above matrix of strength values we can immediately tell that p, for example, is stronger than b according to the β scale by one point, but they are of the same strength according to the α scale (p=α₃β₃; b=α₃β₂).

The notion of strength in the above IDP has been based on several observations made by Foley which indicate that in North German, Danish, and Spanish g spirantises more readily than d or b, and g and d spirantise more readily than b. Foley therefore establishes the above relation ($\frac{g \ d \ b}{1 \ 2 \ 3}$) "which refers to the propensity to spirantization, with the weakest element being most inclined to spirantization... The relation α is the relative phonological strength of elements appearing phonetically as velars, dentals and labials" (1977:28).

The relation α has been determined by whether or not certain elements weaken. Other observations reveal that g weakens to γ but that d does not weaken to ð. This kind of weakening reveals that g is phonologically weaker than d and that γ is weaker than g. These relations ($\frac{2 \uparrow \ g \ d \ b}{1 \ \gamma \ \dot{\delta} \ \beta}$) are called β phonological strength, "which reflects the fact that the weakened reflexes of voiced stops are voiced continuants" (1977:33).

Based on the same observations on Norwegian, Lithuanian, English and Latin, Foley has also created the relative phonological strength ρ as follows:

(2)	t	s	n	l	w	e
	→					
	1	2	3	4	5	6

where t stands for oral stops, s for continuants, n for nasals, l for liquids, w for glides, and e for vowels, and where higher number represents greater relative strength.

3.0. Unfortunately the above inertial development principles do not work the way they are supposed to, even in Germanic languages. According to the IDP, strengthening applies first to the strongest elements. Thus, in the second Germanic consonant shift the voiceless stops (β₃), being the strongest elements, strengthened before the voiced stops (β₂). Now let us consider the following examples that Foley gives (1977:107):

(3)	<u>German</u>	<u>English</u>	
	<i>Tür</i>	'door'	d (α ₂)
	<i>Bart</i>	'beard'	b (α ₃)
	<i>Grab</i>	'grave'	g (α ₁)

Foley states that when the voiced stops do shift to the stronger voiceless stops, only the strongest voiced stop undergoes the shift, while the weaker voiced stops do not undergo the shift. So, in (3) b ($\alpha 3$) is the strongest element according to (1) but it does not shift, while the less stronger d ($\alpha 2$) does.

Another point which shows the cryptotaxonomic¹ nature of this Foleyan framework is its application to handle acoustically caused changes (Ohala 1974). According to Foley (1973:51), his concept of relative phonological strength (IDP) must be included within a theoretical system of phonology to account for what Foley considers as the phonological type of assimilation, that is, the kind of assimilation which cannot be explained phonetically. To clarify his point, Foley gives the following examples in Norwegian:

- | | | | | |
|-----|-----|--------|--------|---------|
| (4) | (a) | slem | [šlem] | 'bad' |
| | | sɬ | [šlo] | 'beat' |
| | | Oslo | [ošlo] | 'Oslo' |
| | (b) | snakke | | 'talk' |
| | | snø | | 'snow' |
| | | snar | | 'quick' |

Foley states that "the conversion of s to š cannot be attributed to phonetic influences since l and n are both dentals, and in any case conversion of dental s to palatal š looks more like dissimilation than assimilation". (1973:51). He argues strongly that "there is no phonetic explanation for the palatalization of s before l but not before n" (p.51). So Foley offers a solution which he calls "a phonological explanation" (p.51) by applying the relative phonological strength ρ ((2) above): "l is phonologically stronger than n, and s is strengthened by proximity to l but not by proximity to the relatively weaker n; the strengthened s then manifests itself as š" (1973:51).

One thing that Foley forgets is that although l and n are both dentals they are two quite different segments especially acoustically, the l being (+lateral) and (-nasal) and the n being (-lateral) and (+nasal). Ohala (1974) has noticed the tendency among many phonologists to think of phonetically plausible processes as being primarily those involving articulatory assimilation by neglecting acoustically-caused changes. In the above examples (4), Foley has failed to explain the changes from the acoustic point of view which is phonetic in nature by immediately jumping to the conclusion that there is no phonetic explanation for the changes. Actually, if in (4) s palatalises to š before l but not before n we should not be surprised, it is quite common in assimilation, because s occurs in two quite different environments, namely (5) (a) and (b) below:

- | | | | | | |
|-----|-----|---|--------|-------|--------|
| (5) | (a) | s | _____> | š/ | _____l |
| | (b) | s | _____> | idem/ | _____n |

Hence, the two different environments are responsible for the acoustically, thus phonetically, plausible change. Einar Haugen (1942) has worked on this kind of change in Norwegian (cf. also Ohala 1974). In the cluster sɬ, the l is partially devoiced because of its proximity to s thus yielding [sɬ̥]. The partially devoiced [ɬ̥] is acoustically a fricative, and according to Haugen (1942) it is in fact very much like [š]. This acoustically caused assimilation in the word slem in (4) can be shown as follows:

(6)	/slɛm/	Underlying Form
	/s̥lɛm/	Partial Devoicing
	/s̥slɛm/	Acoustic Assimilation
	/ʃlɛm/	Cluster Simplification
	[ʃlɛm]	Phonetic Form

In the word *snakke* of (4), however, the cluster *sn* does not cause acoustic assimilation because the devoiced [ŋ] does not yield a fricative sounding [ʃ]. We can see clearly that there is no need to create a new theory (the IDP) just to explain a simple acoustic assimilation.

4.0. Let us now turn to the application of the IDP to Malay and Javanese, the first two Austronesian languages chosen by Mashudi (1979) to verify these principles:

(7) *Malay (Bahasa Malaysia)*

	Prefix-stem	Final form	Glosses
1a.	məŋ + kawal	məŋawal	'to guard'
b.	məŋ + gulun	məŋgulun	'to roll something'
2a.	məŋ + tulis	mənules	'to write'
b.	məŋ + dapat	məndapat	'to get, to obtain'
3a.	məŋ + pukul	məmukul	'to hit'
b.	məŋ + basuh	məmbasuh	'to wash'

Javanese

4a.	ŋ + kidol	ŋidol	'to the south'
b.	ŋ + gole?	ŋgole?	'to look for'
5a.	ŋ + tapo?	napo?	'to slap the mouth'
b.	ŋ + delo?	ndəlo?	'to see'
6a.	ŋ + paku	maku	'to nail'
b.	ŋ + belah	mbəlah	'to split'

In both Malay and Javanese, as in (7) above, Mashudi observes that "t disappears in a strong position (postnasal), while d remains. The disappearance of t in postnasal position is due to postnasal strengthening of t and not to assimilation or elision" (1979:35-36). (The dental t is used by Mashudi as a cover symbol for p and k as well). Thus, applying postnasal strengthening to t first, according to Mashudi, it will yield the correct results for both *mənules* and *məndapat*, as follows:

(8)	məŋ + dapat	məŋ + tules	Underlying Form
	mən + dapat	mən + tules	Ass.I.(ŋ → n/ ___ +d) ___ +t)
	mən + dapat	mən ⁺ tules	Positional Potentiation (t → t ⁺ /N+ ___)
	məndapat	mənules	Modular Depotentiation (t ⁺ → ∅)

In this paper we would like to argue that the disappearance of t in data (7) above is not due to postnasal strengthening, but is simply due to a phonological process called *coalescence of consonants* which is phonetically motivated. Or it can also be solved by the simple assimilation process in the traditional way of Generative Phonology. If it is due to postnasal strengthening t (β3 according to (1)) should not disappear, instead it should be strengthened to tt (β4), and

then we would get the wrong results:

- (9) 1a. məŋ + kawal → məŋ + kkawal → *məkkawal
- 2a. məŋ + tulis → məŋ + ttules → *məttules
- 3a. məŋ + pukul → məŋ + ppukul → *məppukul

The application of the Foleyian 'gadget' of the so-called 'modular depotentiation' by Mashudi to tackle the effacement of t is quite untimely and thus cannot be justified, because after t is strengthened to its strongest form tt it is not strengthened any more and there is no point in effacing it, but instead it should stay in its strongest form. This is also one of the drawbacks of the Foleyian framework of the IDP. When the IDP faces a problem it cannot handle, such as the supposed strengthening of a strong element which results in its lenition, Foley immediately creates a 'gadget' to tackle the problem, that is the 'modular depotentiation', "the special instance of the strongest element on a parameter when strengthened appearing phonetically as the weakest element" (1977:108). While Foley only says "appearing phonetically as the weakest element", Mashudi seems to feel like going a little bit further by effacing the element. This, we think, is a bit strange, because in the first instance, when an element is strengthened it becomes stronger, not weaker, let alone elided.

As has been mentioned above the disappearance of t in (7) can be explained in two ways: (a) by the traditional nasal assimilation process of the Generative Phonology (GP), or (b) by the phonological process of the coalescence of consonants also of the GP. The traditional nasal assimilation (a) is as follows (cf. Farid M. Onn 1976, 1980):

(10)	(a)1	məŋ + dapat	(a)2	məŋ + tulis	Underlying Form
		mən + dapat		mən + tules	Nasal Assimilation
		--		mənules	Cluster Simplification
		məndapat		mənules	Phonetic Form

In (10) (a)1 there is no need to apply the cluster simplification rule, because both n and d have the feature (+voice) which binds/renderers them as an easy-to-produce cluster (unmarked),² while in (10) (a)2 it has to be applied since n and t do not usually go together in Malay (the cluster nt is marked).

The phonological process of the coalescence of consonants (b) is only applied to problem (a)2 of (10) above as in (11) (a)2 below (cf. Simanjuntak 1979):

(11)	(a)1	məŋ + dapat	Underlying Form
		mən + dapat	Nasal Assimilation
		məndapat	Phonetic Form
	(a)2	məŋ + tulis	Underlying Form
		mə-n-ules	Coalescence
		mənules	Phonetic Form

In (11) (a)2 the marked cluster nt undergoes the coalescence process while in (11) (a)1 the unmarked cluster nd does not.

Schane (1973:56) explains the coalescence process as involving "both assimilation and reduction, [it] could be described as the joint action of these two processes". In the above coalescence process both elements ŋ and t must undergo assimilation and reduction in that order : first ŋ must assimilate to t and t to ŋ simultaneously, and then both must be reduced to ø, or one of them

which contains the major features is modified and the other is deleted, to give place to the resultant assimilated form, the intermediary element, that is an element which in between η and t , the choice of which depends on which of the two has the major (stronger) features. In the above case the choice is n .³

In order to decide the intermediary element let us first list the features of both elements, thus:

(12)	Members of the coalescence process		Intermediary element
	η	t	n
voc	-	-	-
ant	-	+	+
cons	+	+	+
cor	-	+	+
hi	+	-	-
vc	+	-	+
bk	+	-	-
cont	-	-	-
lw	-	-	-
nas	+	-	+
str	-	-	-

(Based on Chomsky and Halle 1968)

From (12) above we can immediately see that η has two major features which t lacks, they are [+voice] and [+nasal]. On the other hand t has (hierarchically) high features which η lacks, they are [+anterior] and [+coronal]. In other words η is stronger than t in its voiceness and nasality, but t is stronger than η in its anteriority and coronality. Thus, the intermediary element must be one in between η and t that contains the four types of strength, namely voiceness, nasality, anteriority, and coronality and it must also be [+consonantal], because this feature is contained by both members of the coalescence process. The only element that meets these conditions is n (see (12) above). The above coalescence process which involves assimilation and reduction can be stated by the following rules:

- (13) (a) $\left\langle \begin{array}{c} v\eta \longrightarrow n/ v \quad t \\ tV \longrightarrow n/ \eta \quad v \end{array} \right\rangle$ or
- (b) $\left\langle \begin{array}{c} v\eta \longrightarrow n/ v \quad t \\ tV \longrightarrow \phi/ \eta \quad v \end{array} \right\rangle$

Either rule (13) (a) or rule (13) (b) can be chosen according to one's preference: in rule (13) (a) both members of the coalescence process (η and t) are changed (reduced) to n , while in rule (13) (b) one member (η) is modified to n and the other (t) is deleted (cf. Schane 1973:68).

In the data from Javanese above (7), two processes are responsible for the sound changes: total assimilation and coalescence.

- (14) (a)1 $\eta + kidol$ $\eta idol$ '*to the south*'
 (a)2 $\eta + tapo?$ $napo?$ '*to slap the mouth*'

In (14) (a)1 total assimilation process takes place, that is k is totally assimilated to ŋ, while in (14) (a)2 coalescence process yields the phonetic form [napo?]:



5.0. Let us now turn to Indonesian data. The same processes above can be applied to explain sound changes in the data from Indonesian taken by Mashudi (1979:38) below:

(15) Indonesian

	Prefix-stem	Final form	Glosses
(a)1	məŋ + kawal	məŋawal	'to guard'
(a)2	məŋ + pukul	məmukul	'to hit'

In (15) (a)1 total assimilation process is responsible for the elision of k in [məŋawal], while in (15) (a)2 coalescence process may be used to yield the phonetic form [məmukul]:



In this case the strength of the element p lies in its anteriority only, so the intermediary element should contain the feature [+anterior] in addition to the other two major features belonging to ŋ, which are [+voice] and [+nasal] and the shared feature [+consonantal], so the intermediary element is m (see (16) below):

(16) Members of the coalescence process Intermediary element

	ŋ	p	m
voc	-	-	-
ant	-	+	+
cons	+	+	+
cor	-	-	-
hi	+	-	-
vc	+	-	+
bk	+	-	-
cont	-	-	-
lw	-	-	-
nas	+	-	+
str	-	-	-

(Based on Chomsky and Halle 1968)

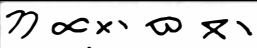
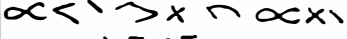


6.0. Let us now observe the following data from Toba Batak⁴ taken by Mashudi (1979:39) and compare them to the Proto-Austronesian (PAN) forms:

(17)	PAN	Toba Batak	Glosses
(a)	*tunked	tukkot	'a walking stick'
(b)	*zempu	ʝopput	'to pick up with the hand'
(c)	*dembar	hombar	'twin'
(d)	*sumpel	suppol	'a cork'
(e)	*pentas	pottas	'a platform'

In the above data (17) Mashudi refers to the development of *ŋk > kk in (17) (a), *mp > pp in (17) (b) and *nt > tt in (17) (e), but *mb > idem in (17) (c). Mashudi admits that he does not have examples for *nd > idem and *ng > idem. However, he takes these forms (in orthography): manggolom, mandege, and mambuwat, as additional data to support his argument, that in postnasal position in Toba Batak g, d, b, do not undergo strengthening and this in turn lends support to his claim that the IDP is universally valid. Unfortunately, in actual pronunciation, (17) (c) and the additional data are not as truthful as they look orthographically, because they are actually represented phonetically as follows (thus, g, d, b also undergo strengthening if the so-called strengthening really takes place):

(18)	hombar	[hobbar]	'twin' or 'next to'
	manggolom	[maggolom] ⁵	'to hold in the hand'
	mandege	[maddege]	'to set foot on'
	mambuwat	[mabbuwat]	'to take'

Of course Mashudi must not be blamed for this since Toba Batak data are usually taken from Van der Tuuk's books especially since one of them has been translated into English. Dempwolff (1934-38) and Dyen (1968) and others also take Toba Batak data from Van der Tuuk's books. Toba Batak is a notoriously difficult language to transliterate, and all the materials, especially the Batak 'pustaha' (manuscripts) studied and consulted by Van der Tuuk were written in Toba Batak script. In Toba Batak script the data in (18) are represented as follows (cf. Van der Tuuk 1862, 1971; Sibarani 1970):

(19)	Original form	Transliterated form
		hombar
		manggolom
		mandege
		mambuwat

The transliterated form is not the actual phonetic form of a word but it is the underlying form. So the actual phonetic representations of (19) are as in (18).

The translator of Van der Tuuk's *Tobasche spraak(k)unst* (A grammar of Toba Batak 1971) has also erred linguistically (or rather phonetically) speaking for putting all the Toba Batak transliterated forms in square brackets as if they were the actual forms in pronunciation.

From the contradictory data in (18) and (19) an immediate conclusion can be drawn that the IDP is inadequate as has been implicitly mentioned by Mashudi when he says: "I do not have examples for *nd > idem and *ng > idem" (1979:39).

The examples do not exist, however see note 4 in the case of *ŋg > idem, when the word is pronounced very slowly.

To clarify our point let us consider the following data from Toba Batak (Van der Tuuk 1971):

(20)	Transliterated form (Underlying form)	Phonetic form	Glosses
(a)	/tintinhu/	[tittittu] or [tittikku]	'my ring'
(b)	/napuranhu/	[napurattu] or [napurakku]	'my betel'
(c)	/onomhalak/	[onoppalak] or [onokkalak]	'six people'
(d)	/onomhorbo/	[onopporbo] or [onokkorbo]	'six buffaloes'

We think that the IDP is quite inadequate to explain the sound changes in data (20) above. However, let us try to apply the IDP to account for the development of nt > tt and nh > tt or kk in (20) (a) following exactly the steps Mashudi takes to explain the sound changes in data (17):

(21)	tintinhu	Underlying Form
	tint ⁺ inh ⁺ u	Positional Potentiation
	tinttinhhu	Depotentiation (t ⁺ → tt) (h ⁺ → hh)
	*tittihhu	Nasal Effacement (N → φ / ___ CC)

Here we get the final form +tittihhu which is not the correct form. If h (ρ2 according to scale (2)) is strengthened because of its proximity to the stronger n (ρ3 according to scale (2)), does it become h⁺ or k or what? Suppose it becomes k and not h⁺ like in (21), then the process will proceed like (22) below:

(22)	tintinhu	Underlying Form
	tint ⁺ inku	Positional Potentiation
	tinttin?u	Depotentiation (t ⁺ → tt) (k → ? according to the αβ scale)
	*titti?u	Nasal Effacement (N → φ / ___ CC) (N → φ / ___ C)

Actually when we come to the third step we get stuck because we do not have any cogent reason to depotentiate t⁺ in the cluster nt⁺ and k in the cluster nk. If, however, k should be depotentiated for no reason at all then we would get ? according to the combination of the α and the β scales. Then after the nasal effacement we get the wrong form: *[titti?u].

Another thing which is quite puzzling is the use of t⁺ and tt in the IDP. First Foley (1970) uses t⁺ (β4) to indicate the strengthened t (β3) and this 1970 version is used by Mashudi (1979). But then later Foley (1977) uses tt (also β4) to indicate the strengthened t (β3). From the use of the two versions we may conclude that t⁺ is the same as tt. But then Mashudi uses tt to indicate that

t^+ has been depotentiated, although the reason for this depotentiation is unclear. Foley uses or creates the so-called 'modular depotentiation' rule only when "the element is already the strongest element and cannot appear phonetically as a stronger element. In this case,... the strengthened strongest element undergoes modular depotentiation, appearing phonetically as the weakest element" (1977:123). In the above example (22) there is no reason to strengthen t twice. The first strengthening is reasonable because of its proximity to the stronger element n , but the second strengthening is baseless. If for no reason whatsoever the element t^+ must be strengthened then it should under the modular depotentiation rule because t^+ is already in its strongest form (β_4). If t^+ undergoes modular depotentiation then it will become the weakest element δ according to the β scale or k according to the α scale. In Toba Batak the element δ does not occur, so t^+ must undergo total weakening, that is, it must be effaced according to Foley (1977:123). So t^+ after strengthening should become either \emptyset or k : whichever it is the final form is wrong. We can see how inadequate the IDP is in handling complicated sound changes of this nature.

According to our observations two phonological processes, assimilation (total or partial) and coalescence, are quite predominant in Austronesian languages. From our point of view (20) and (21) can be solved as follows:

(23) tintinhu	Underlying Form
nt → tt	Total Regressive Assimilation (n is totally assimilated to t and itself becomes t)
nh → tt or → kk	Coalescence ($\langle \begin{smallmatrix} n & \rightarrow & t \\ h & \rightarrow & t \end{smallmatrix} \rangle$ or 6) $\langle \begin{smallmatrix} n & \rightarrow & k \\ n & \rightarrow & k \end{smallmatrix} \rangle$)
tittittu or tittikku	Final Form

The coalescence process in (23) deserves further explanation. Both members of the process change to the same intermediary element: to either t or k depending on which member is preferred. If member n is preferred then the intermediary element must be a consonant which is homorganic with n , that is t ; however if the preferred member is h , then the intermediary element must be a consonant which is homorganic with h , that is k .

Problem (20) (b) can be solved in the same way as problem (20) (a), that is by solution (23). Problems (20) (c), (d), (e) can be solved as follows:

(24) mh → pp or → kk	Coalescence ($\langle \begin{smallmatrix} m & \rightarrow & p \\ h & \rightarrow & p \end{smallmatrix} \rangle$ or $\langle \begin{smallmatrix} m & \rightarrow & k \\ h & \rightarrow & k \end{smallmatrix} \rangle$)
	The above can be stated as below:

$$\left\{ \begin{smallmatrix} m \\ h \end{smallmatrix} \right\} \rightarrow \left\{ \begin{smallmatrix} p \\ k \end{smallmatrix} \right\} / v \text{ ___ } v$$

We can see that in (24) the intermediary element is either p or k ; the first preference is p which is homorganic with m and the second preference is k which is homorganic with h .

7.0. Let us now turn to Samoan data below given by Mashudi to support the universality of the IDP:

(25)	PAN	Samoan	Glosses
(a)	*pulog	fulu	'ten'
(b)	*panaq	fana	'shoot'
(c)	*xepat	fa	'four'
(d)	*apuy	afi	'fire'
(e)	*taqun	tau	'year'
(f)	*tanis (*caŋ is according to Dyen 1971)	tani (tani, Dyen 1971)	'cry'
(g)	*mata	mata	'eye'
(h)	*kaw	?au	'you'
(i)	*kutuh	?utu	'louse'
(j)	*ikan	i?a	'fish'
(k)	*aku	a?u	'I'

In data (25) above Mashudi is concerned with the shift from *p > f, while *t > idem and *k > ?. To these problems he combines the $\alpha\beta$ parameters at $\beta:3$, and he comes up with the following scale:

$$(26) \quad \begin{array}{ccccccc} ? & k & t & p & f & & \\ \hline & & & & & & \rightarrow \\ 4 & 5 & 6 & 7 & 8 & & \end{array}$$

According to scale (26), p with a strength value of 7 is stronger than either t or k. He concludes that in Samoan PAN *p undergoes strengthening to f to yield (25) (a) ___ (d) in initial and syllable onset positions. However, in (e) ___ (g) t being weaker than p, although stronger than k, does not strengthen, while k with $\alpha\beta:5$ weakens to ? as in (25) (h) ___ (k).

We are not quite sure how Mashudi figures out the scale above, that is (26), because according to the $\alpha\beta$ scale (1), if we combine the $\alpha\beta$ parameters at $\beta:3$ we will have scale (27) below:

$$(27) \quad \begin{array}{ccccccc} ? & k & t & p & f & & \\ \hline & & & & & & \rightarrow \\ 3 & 4 & 5 & 6 & 7 & & \end{array}$$

Thus p is $\alpha_3 \beta_3:6$ and not $\alpha_4 \beta_3:7$, t is $\alpha_2 \beta_3:5$ not $\alpha_3 \beta_3$, and k is $\alpha_1 \beta_3:4$ not $\alpha_2 \beta_3:5$, and if we include g and b we will have scale (28) below:

$$(28) \quad \begin{array}{ccccccc} & g & & b & & & \\ & ? & k & t & p & f & \\ \hline & & & & & & \rightarrow \\ 3 & 4 & 5 & 6 & 7 & & \end{array}$$

where g ($\alpha_1 \beta_2:3$) and ? have the lowest strength value, while b ($\alpha_3 \beta_2:5$) and t ($\alpha_2 \beta_3:5$) have the same strength value. Bearing in mind scale (28), let us consider the following data:

(29)	PAN	Samoan	Glosses
(a)	*gilan	ki-kila	'glitter, sparkle'
(b)	*belaj	fola	'spread out'
(c)	*ubi	ufi	'yam'

In data (29) (a) *g which is the lowest in terms of strength value according to scale (28) is strengthened to k which is against the IDP which says that weak elements weaken first and most extensively and preferentially and strong elements strengthen first and most extensively and preferentially in strong environments. As we have seen in data (25) above, k (strength value 4) and t (strength value 5) do not undergo strengthening, but in data (29) g (strength value 3) as the weakest of them all does. Even though k is stronger than g, it nevertheless undergoes weakening to ?. Another case is b (strength value 5) in (29) (b) and (c) which undergoes double strengthening to f (strength value 7) while the strongest element p (strength value 6) undergoes single strengthening only, and t which has the same strength as b does not undergo strengthening at all.

Considering the contradictory data above we strongly feel that the IDP is quite inadequate to account for sound changes in Samoan and in other Austronesian languages.

8.0. Let us now turn to Tagalog data in which, according to Mashudi, b strengthens while d and g do not:

- (30) (a) maŋ + guló > maŋguló 'to make trouble'
 (b) ipaŋ + damó > ipandamó 'to weed'
 (c) maŋ + bilí > mamilí 'to buy'

In data (30) Mashudi argues that "while b strengthens in a strong position and then returns to the lowest scale, g and d remain neutral in that position" (1979:41). Then Mashudi provides more data to support his argument. He says: "Since b strengthens in postnasal position, according to the IDP, k, t, p, should also strengthen. That they behave as expected may be illustrated by the following examples" (1979:4):

- (31) (a) maŋ + tahí > manahí 'to sew'
 (b) maŋ + kúha > maŋúha 'to get'
 (c) maŋ + píli? > mamíli? 'to choose'

Based on data (30) and (31) Mashudi draws the following conclusion: "Therefore, in agreement with the IDP, the Tagalog data have illustrated that strengthening applies to strong elements, namely k, t, p, b in a postnasal position; but d and g do not strengthen in that position" (1979:42).

If we study data (30) and (31) closely and nothing else we will be tempted immediately to believe that the IDP is indeed quite adequate to account for the above sound changes. But if we look at some other data from Tagalog we will be persuaded to question the adequacy of the IDP. Let us look at the following data (cf. Panganiban 1972; Aspillera 1974) and compare them with data (30) and (31):

- (32) cf. data (30) (a):
- | | | |
|-----|-------------------------------------|---|
| (a) | maŋ + guló → manguló
→ maguló | 'to cause trouble'
'troublesome' |
| (b) | maŋ + gulay → magulay
→ maggulay | 'having plenty of vegetables'
'to cook and eat vegetables' |
| (c) | maŋ + gusót → magusót | 'to become crumpled' |
| (d) | maŋ + gutom → magutom | 'to be hungry' |
| (e) | maŋ + galit → maǵalit | 'to be angry' |
- (33) cf. data (30) (b):
- | | | |
|-----|---------------------------|------------------------------|
| (a) | maŋ + dalanin → manalanin | 'to pray' |
| (b) | maŋ + dalasa → manalasa | 'to assault' |
| (c) | maŋ + dalapán → manalapán | 'to ride in a large rowboat' |
| (d) | maŋ + dikít → manikít | 'to adhere' |
- (34) cf. data (30) (c):
- | | | |
|-----|--|---|
| (a) | maŋ + bilí → manilí
→ mabilí
→ magbilí
(bumilí : 'to buy')
(infix: -um-) | 'to make purchases'
'to be able to buy'
'to sell' |
| (b) | maŋ + bihasa → manihasa
→ mabihasa | 'to get used to a certain practice'
'to get used to' |
- (35) cf. (31) (a):
- | | | |
|-----|--|--|
| (a) | maŋ + tahan → magtahan | 'to stop' |
| (b) | maŋ + tahip → magtahip | 'to winnow grains' |
| (c) | maŋ + talo → matalo
→ manalo
→ magtalo | 'to lose'
'to win'
'to argue with each other' |
| (d) | maŋ + tanán → magtanán | 'to escape' |
| (e) | maŋ + tanón → magtanón
→ matanón | 'to ask about something'
'to happen to be asked a question' |
- (36) cf. (31) (b):
- | | | |
|-----|---|--|
| (a) | maŋ + kuha → maŋuha
→ makuha
→ mákuha | 'to gather, to collect'
'to be able to get'
'to happen to get' |
| (b) | maŋ + kulam → makulam
→ mákulam | 'to be able to bewitch something'
'to become bewitched' |
| (c) | maŋ + kulaŋ → magkulaŋ | 'to fail to please due to
shortcomings' |
| (d) | maŋ + kulambô → magkulambô | 'to use a mosquito net' |
| (e) | maŋ + kurtina → magkurtina | 'to use curtains' |
| (f) | maŋ + kusa → magkusa | 'to do a thing voluntarily' |
| (g) | maŋ + kuskós → magkuskós | 'to scrub' |

(37) cf. (31) (c):

- | | | |
|-----|--|--|
| (a) | maŋ + píli → mamíli
→ mapíli | 'to choose'
'choosy' |
| (b) | maŋ + pilay → mapilay | 'to become tame' |
| (c) | maŋ + piliŋ → magpiliŋ | 'to be close together' |
| (d) | maŋ + pilipit → mamilipit
→ mapilipit | 'to become twisted'
'to be able to twist something' |
| (e) | maŋ + pilit → mamilit
→ magpilit | 'to force'
'to try hard' |
| (f) | maŋ + pinsala → maminsala
→ mapinsala | 'to injure'
'to be injured' |
| (g) | maŋ + piraso → mapiraso | 'to be broken into pieces' |
| (h) | maŋ + piriŋ → magpiriŋ | 'to cover one's eyes with a blind-fold' |

Data (32)-(37) above are enough to tempt us to hypothesise that in Tagalog the underlying form /maŋ/ is phonetically realised in different forms ([maŋ], [man], [ma], [mag]) depending on the meaning of the word in which it is used. In other words the phonetic representations of the underlying form /maŋ/ seem to be mostly semantically motivated although in some cases phonological motivations also try to manifest themselves. However, we can also hypothesise that in Tagalog there are four underlying forms (/maŋ/, /man/, /ma/, /mag/) which manifest themselves in the same forms phonetically to form different types of verbs, thus as [maŋ], [man], [ma], [mag]. In data (32)-(37) above the first hypothesis has been applied.

By applying the first hypothesis, we think that the IDP is inadequate to explain the sound changes that Tagalog manipulates to convey different meanings. In data (32), for example, ŋ (ρ3) should not disappear from the cluster ŋg according to the IDP because it is stronger than g (ρ1) according to the ρ scale (2). In data (33), d should not disappear according to the IDP (cf. data (30) (b)), because it must not strengthen in postnasal position due to its weaker strength value (α2 β2). In data (34) [mamiŋí] and [mamihasa] are the correct forms according to the IDP, but we also have [mabiŋí], [magbiŋí], and [mabihasa]. And so on and so forth.

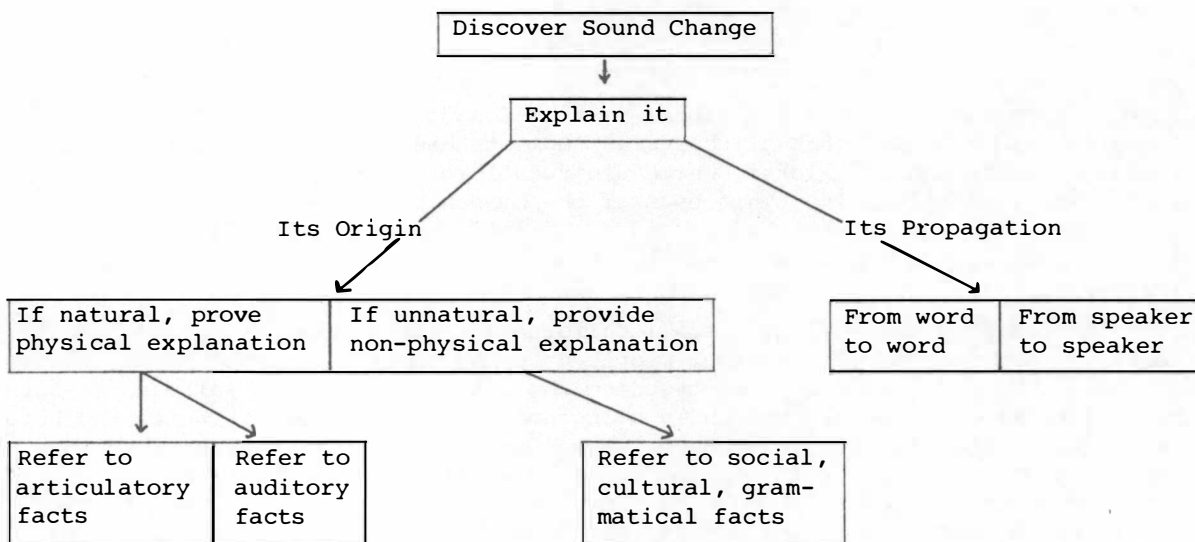
From the above data we can conclude that Tagalog is a language that manipulates assimilation and coalescence processes not only for the sake of ease of production but most importantly for distinguishing different meanings. These kinds of phonological processes are said to be semantically motivated in the sense that phonological processes are used or are made good use of for semantic purposes. As an illustration let us take just one example: data (34) (a):

- | | | | | |
|------|-----|--------------|-----------------|-------------------------------|
| (38) | (a) | bilí | Root | 'the amount paid for a thing' |
| | | /maŋ + bili/ | Underlying Form | |
| | | ↙ ↘ | | |
| | | m | | |
| | | mamili | Coalescence | < ŋ → m / b → m > |
| | | [mamiliŋ] | Phonetic Form | 'to make purchases' |

- (b) /maŋ + bili/ Underlying Form
 mabbili Total Assimilation n → b/ ___ b
 mabili Cluster Simplification
 [mabilí] Phonetic Form 'to be able to buy'
- (c) /maŋ + bili/ Underlying Form
 magbili Partial Assimilation N → Stop/ ___ Stop
 [magbilí] Phonetic Form 'to sell'

(Note: In order to have a word with the meaning 'to buy', the language uses the infix um (-um-), thus we have the word bumilí with that meaning.)

9.0. So far, Samoan is the only language for which we have not given alternative solutions to its problems of sound changes. This is because of the fact that we consider sound changes in Samoan to be unnatural to the extent that we cannot find articulatory or auditory (acoustic) facts or even semantic motivations to account for these sound changes. To show the unnaturalness of the Samoan sound changes, let us consider the consonant segments of this language. Samoan has the following consonant segments: p t ʔ m n ŋ f s v l (Biggs 1971: cf. Churchward 1951). Two segments h and r are only used in words of foreign origin, and k is used only in foreign words (e.g. saka 'to boil' and puka 'an interjection' are words from some neighbouring islands). Another unusual thing about the phonology of this language is that in colloquial Samoan the segment n has changed to ŋ, a fact which contradicts the statement made by Ferguson in Greenberg (1966:57) that if there are only two nasals they will be m and n. The segment p also is very rare and it usually occurs in loanwords. Because of the unusualness of the Samoan segments, we do not think that the IDP is adequate to explain this fact. So we have to refer to other facts, non-physical facts (cf. J. Ohala 1974b: the diagram below), and here we would suggest that we refer to (consider) sociolinguistic as well as psycholinguistic aspects.



We would like to argue here that these non-physical facts are just as important and relevant as the physical (neurolinguistic) facts in accounting for sound changes in natural languages (cf. J. Ohala 1974b; Slobin 1977; Baron 1977; Goyvaerts 1975; Andersen 1973; Simanjuntak and Ramli Salleh 1980). These non-physical factors are unique to a particular language, so they must be language-specific and culture-specific (J. Ohala 1974b). Some of the non-physical factors that Malkiel mentions (1973: see also J. Ohala 1974b) are: hypercorrection, fashionable pronunciations (cf. Postal 1968 who suggests that sounds change due to the same whims of fashion that cause clothing and car styling to change), spelling pronunciation, analogical extension of grammatical paradigm. Here we would like to report a unique observation made by Simanjuntak (1971) which illustrates how a sociolinguistic as well as psycholinguistic incident causes sound changes in the language of children under seven years of age.

One day, in a remote village in Tapanuli (North Sumatra), Simanjuntak overheard small children (under seven years of age) being taught by older children how to pronounce correctly the numbers from one to ten in Toba Batak, the mother tongue of the children. The small children's pronunciation was strange which made the older children laugh and so they tried to teach them the correct pronunciation. This is how the small children counted from one to ten:

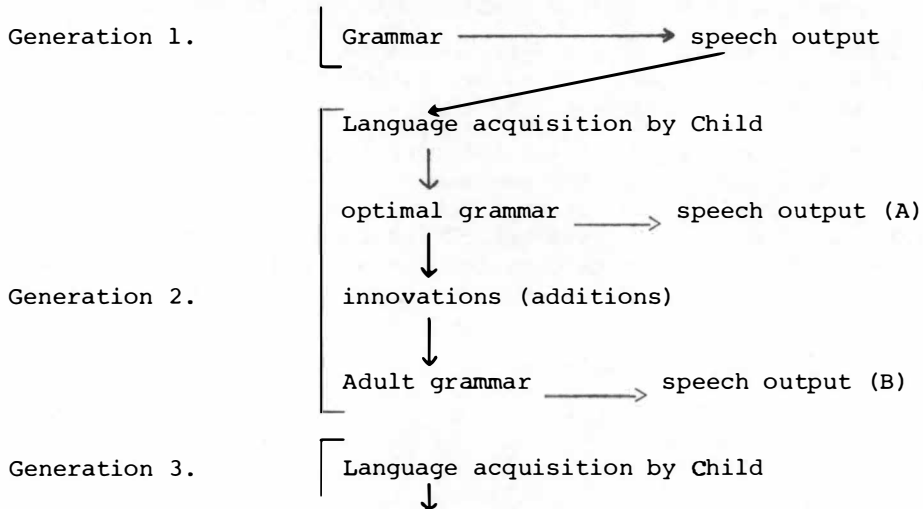
Children's Pronunciation	(Correct) Adult's Pronunciation	
[sada]	[sada]	'one'
[dua]	[dua]	'two'
[tolu]	[tolu]	'three'
[okat]	[opat]	'four'
[liŋa]	[lima]	'five'
[onoŋ]	[onom]	'six'
[kitu]	[pitu]	'seven'
[ualu]	[ualu]	'eight'
[sia]	[sia]	'nine'
[sakkulu]	[sappulu]	'ten'

What is apparent is that all bilabial segments changed to velars. Further investigation revealed that all bilabial segments had indeed disappeared from the language of the small children, while older children seemed to have managed to regain the lost bilabials. This unusual phenomenon had urged the investigator to do further study as to what had caused the disappearance of the bilabials from the language of the small children: the observed situation appears to be as follows.

In that particular village small children are taken care of by older people, usually grandmothers. These older people have the habit of chewing betel leaves and tobacco and at the same time teaching their grandchildren to talk and to count by telling stories and all the while doing small household tasks. While the betel leaves are being chewed, the tobacco 'clump' or 'ball' is inserted between the upper lip and the upper teeth or gum creating a situation in which the upper lip can never meet the lower lip when talking. The situation is aggravated by the betel leaves which are placed between the upper and lower molars preventing the upper teeth from reaching the lower lip. We can imagine how impossible it is for these eager grandmothers to produce bilabial sounds when teaching their grandchildren, thus the easiest way for them, which is quite natural, is to replace

all bilabials with velars. To make the situation worse, the younger adults have the habit of imitating the nonbilabial-sound language of the grandmothers to create humorous situations in the presence of the small children. Thus, practically speaking, the small children are acquiring this nonbilabial-sound language from the start. However, when the children start schooling their defective language is corrected by the teachers because Toba Batak is the medium of instruction in the elementary school up to standard three (beginning standard four, Bahasa Indonesia, the national language, takes over as the medium).

The above incident is a good example of the many possible similar incidents that might (or must) have taken place during the many thousands of years which separate Samoan from its Proto-Austronesian ancestor and which have resulted in the present disappearance of b, h, r, k from its phonemic inventory. These incidents must have caused the language to restructure its phonological grammar many times which must have resulted in the repeated additions, losses, and reorderings of its phonological rules (cf. King 1969; Kiparsky 1965, 1968; Farid Onn 1980) or probably inversions of these rules (cf. Vennemann 1972). Non-physical causes of sound changes are indeed very difficult to describe precisely as to how they have caused a particular language to change its sounds. We can only guess. But roughly the sociolinguistic and psycholinguistic processes of sound changes can be illustrated by the following diagram (Goyvaerts 1975:128):



(Linguistic change is change in competence not in performance.)

NOTES

1. It appears that the IDP is also inadequate in handling the European languages, hence J. Ohala (1974) may be justified in saying that Foley has cryptotaxonomic goals with this theory of phonology, which Foley claims to be "perhaps the only genuine theory of phonology in existence" (1971:1).
2. In Ulu Muar and Kedah (Langkawi) dialects (both are Malay dialects), however, this cluster simplification rule takes place, so that the cluster *nd* is simplified to *n*, although both elements have the feature [+voice], thus /*men + daki*/ → [mə*n*aki], /*men + dapat*/ → [mə*n*apat] (see Hendon 1966 for Ulu Muar and Collins 1976 for Kedah; cf. Farid M. Onn 1976, 1980a).
3. In traditional descriptive linguistics this coalescence process is known as internal sandhi, a term borrowed from Sanskrit meaning a 'putting together' (in this case) of two consonants. However, how this 'putting together' is actually motivated to produce a third consonant is not explained.
4. Toba Batak is one of the six Batak dialects (the others are Mandailing, Karo, Angkola, Simalungun, Dairi or Pak-Pak) spoken in North Sumatra, Indonesia. The Batak language belongs to the Austronesian family (Simanjuntak 1976).
5. But according to Siahaan (1975) the consonant *n* does not change if the following consonant is *g*, for example in the word *tanggurung* [tang*u*ruŋ]. This may be true in slow speech especially when the word is pronounced in isolation. However, in rapid speech and when the word is used in a sentence, such as *Tingki manggarut tanggurunhu*, the *ŋ* changes to *g*: [tikki maggarut tagguru*g*ku]. This has been tested by Simanjuntak with four native speakers of Toba Batak.
6. These can be stated as follows:

$$\left\{ \begin{matrix} n \\ h \end{matrix} \right\} \rightarrow \left\{ \begin{matrix} t \\ k \end{matrix} \right\} /v _ v$$

BIBLIOGRAPHY

ANDERSEN, Henning

1973 Abductive and deductive change. *Language* 49/4:765-793.

ASPILLERA, P.S.

1974 *Lessons in basic Tagalog for foreigners and non-Tagalogs*. Manila.

BARON, Naomi S.

1977 *Language acquisition and historical change*. Amsterdam: North-Holland.

BIGGS, Bruce

1971 The languages of Polynesia. In: T.A. Sebeok, ed. *Current trends in linguistics*, vol.8: *Linguistics in Oceania*, 466-505. The Hague: Mouton.

CHOMSKY, Noam and Morris HALLE

1968 *The sound pattern of English*. New York: Harper and Row.

CHURCHWARD, Spencer

1951 *A Samoan grammar*. 2nd edn. Melbourne: Methodist Church of Australia.

COLLINS, James T.

1976 Vokal sengau di dalam Bahasa Melayu Kedah (Langkawi). *Dewan Bahasa* 20:19-31.

DEMPWOLFF, Otto

1934-38 *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols. *ZES* 15(1934); *ZES* 17(1937); *ZES* 19(1938). Berlin: Reimer.

DYEN, I.

1965 *A lexicostatistical classification of the Austronesian languages*. *IUPAL Memoir* 19, supplement to *International Journal of American Linguistics* 16.

1971 The Austronesian languages and Proto-Austronesian. In: T.A. Sebeok, ed. *Current trends in linguistics*, vol.8: *Linguistics in Oceania*, 5-54. The Hague: Mouton.

FARID M. ONN

1976 Aspects of Malay phonology and morphology. Ph.D. thesis, University of Illinois, Urbana.

1980a Perubahan bahasa dan kajian dialek: suatu pendekatan tatabahasa generatif. *Dewan Bahasa* 24/8:4-12.

1980b On Grimm's conception of sound-shift and generative phonology. *Akademika* 16:1-11.

FOLEY, James

- 1970 A systematic phonological interpretation of the Germanic consonant shifts. *Language Sciences* 9:11-12.
- 1973 Assimilation of phonological strength in Germanic. In: S.R. Anderson and P. Kiparsky, eds *A Festschrift for Morris Halle*, 51-58. New York: Holt, Rinehart and Winston.
- 1977 *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.

GOYVAERTS, Didier L.

- 1975 *Present-day historical and comparative linguistics: an introductory guide to theory and method*, Part 1: *General background: phonological change*. Ghent-Antwerp: E. Story-Scientia.

GREENBERG, J.H., ed.

- 1966 *Universals of language*. 2nd edn. Cambridge, Mass.: The M.I.T. Press.

HALLE, Morris

- 1962 Phonology in generative grammar. *Word* 18/1:54-72.

HAUGEN, Einar

- 1942 Analysis of a sound group: s' and t' in Norwegian. *PMLA* 57:879-907.

HENDON, Rufus S.

- 1966 *The phonology and morphology of Ulu Muar Malay*. Yale Publications in Anthropology 70. New Haven: Yale University.

KING, Robert D.

- 1969 *Historical linguistics and generative grammar*. Englewood Cliffs, N.J.: Prentice-Hall.

KIPARSKY, Paul

- 1965/ *Phonological change*. (Ph.D. dissertation, M.I.T., Cambridge, Mass.)
1971 Bloomington: Indiana Linguistics Club.
- 1968 Linguistic universals and linguistic change. In: E. Bach and R.T. Harms, eds *Universals in linguistic theory*, 170-202. New York: Holt, Rinehart and Winston.

MALKIEL, Yakov

- 1973 Phonological irregularity vs. lexical complexity in diachronic projection... In: B.B. Kachru et al., eds *Issues in linguistics: papers in honor of Henry and Renée Kahane*, 606-635. Urbana: University of Illinois Press.

MASHUDI B.H. KADER

- 1979 The inertial development principles in phonology: a verification from Malay, Indonesian, Javanese, Toba Batak, Tagalog and Samoa. *Jurnal Budaya Melayu* 4/1:34-56.

OHALA, John J.

- 1974a Phonetic explanation in phonology. In: A. Bruck et al., eds *Papers from the Parasession on natural phonology*, 251-274. Chicago: Chicago Linguistic Society.

OHALA, John J.

- 1974b Experimental historical phonology. In: J.M. Anderson and C. Jones, eds *Historical linguistics*, vol.2: *Theory and description in phonology*, 353-389. Proceedings of the First International Conference on Historical Linguistics, Edinburgh. Amsterdam: North-Holland.

PANGANIBAN, José Villa

- 1972 *Diksyunaryo-tesauro Pilipino-Ingles*. Lungsod ng Quezon: Manlapaz.

POSTAL, Paul M.

- 1968 *Aspects of phonological theory*. New York: Harper & Row.

SCHANE, Sanford A.

- 1973 *Generative phonology*. Englewood Cliffs, N.J.: Prentice-Hall.

SIAHAAN, Nalom

- 1975 *Morfologi bahasa Batak Toba*. Jakarta: Fakultas Sastra, Universitas Indonesia.

SIBARANI, A.N.P.

- 1970 *Buku parsiajaran surat Batak*. Pematang Siantar: Parda.

SIMANJUNTAK, Mangantar

- 1971 Laporan peninjauan di uluan mengenai perubahan Bahasa Batak Toba. FKSS, IKIP, Medan (unpublished report).
- 1976 The Batak script as an invention of the Austronesian-speaking people. *Akademika* 9:59-76.
- 1979 Proses persenyawaan dalam fonologi bahasa Melayu. Universiti Kebangsaan Malaysia. MS.

SIMANJUNTAK, Mangantar and Ramli SALLEH

- 1980 Perubahan-perubahan bunyi bahasa: satu kajian kes Bahasa Melayu Melaka. *Dewan Bahasa* 10:31-43.

SLOBIN, Dan I.

- 1977 Language change in childhood and in history. In: J. Macnamara, ed. *Language learning and thought*, 185-214. New York: Academic Press.

TAYLOR, Mary V.

- 1974 The great Southern Scots conspiracy: pattern in the development of Northern English. In: J.M. Anderson and C. Jones, eds *Historical linguistics*, vol.2: *Theory and description in phonology*, 403-426. Proceedings of the First International Conference on Historical Linguistics, Edinburgh. Amsterdam: North-Holland.

TUUK, H.N. van der

- 1862 *Bataksche Leesboek*. Amsterdam: Frederik Muller.
- 1864/ *A grammar of Toba Batak* (translated by J. Scott-Kemball). The Hague:
1971 Martinus Nyhoff.

VENNEMANN, Theo

- 1972 Rule inversion. *Lingua* 29:209-242.

WEINREICH, Uriel, William LABOV, and Marvin I. HERZOG

- 1968 Empirical foundations for a theory of language change. In: W.P. Lehmann and Y. Malkiel, eds *Directions for historical linguistics: a symposium*, 97-195. Austin: University of Texas Press.

PHONOLOGICAL PROBLEMS OF LOANWORDS IN BAHASA INDONESIA

Hans Lapoliwa

1. INTRODUCTION

Borrowing is quite common in living languages throughout the world. No matter how 'modern' a language is, it must have borrowed certain words or other linguistic elements from other languages in the course of its history, and will continue to do so, for no language can ever be so perfect as to have all the necessary words required to talk about any possible topic in the past, at present, or in the future. The difference between a modern language and a less modern one in this respect lies, probably, only in the degree of borrowing they may exhibit at the same period of history. The latter tends to borrow more items than the former does. Furthermore, a modern language – especially one recognised as an international language – tends to serve as a source language for a less modern one in its efforts to achieve a modern stage through borrowing.

When a foreign word is introduced into a language, it tends to undergo some modification. The degree of such modification depends largely upon the well-formedness conditions of words in the language concerned. The more the foreign item deviates from the well-formedness conditions, the greater the degree of possible modification it tends to undergo. Failure to conform to the well-formedness conditions will cause certain problems to the users of the language concerned, and, to a lesser degree, to the linguist who searches for linguistically significant generalisations (cf. Chomsky and Halle 1968:296).

This paper tries to account for a number of phonological problems which resulted from the introduction of a number of non-Malay words into Bahasa Indonesia, either from foreign languages or from regional languages.¹ Specifically, this paper tries to account for loanwords which are phonologically or morphologically unusual compared with the majority of Bahasa Indonesia words (or Malay words) which cause diversities in pronunciation and/or spelling among the users of Bahasa Indonesia. Quite a number of the diversities seem to be explainable from the linguistic point of view. In this connection, discussion in the subsequent paragraphs will concentrate on the problem of diversities vis-a-vis well-formedness in Bahasa Indonesia of Malay origin. The examples related to pronunciation presented in this paper are based on the speech of a number of people of different background that I managed to observe², while the examples related to spelling are based upon a number of materials written after 1972, the year of the introduction of the present spelling system of Bahasa Indonesia. The main source however, is Poerwadarminta's dictionary.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 285-297. *Pacific Linguistics*, C-75, 1982.

© Hans Lapoliwa

It must be noted from the outset that the study reported here is descriptive in nature. Therefore, no attempt is made to present a complete picture of all existing loanwords, let alone an account of their etymology. The problems related to suprasegmental features are not less interesting but have been intentionally avoided. In brief, loanwords of European origin seem to be marked by stress placed in the last syllable.

It will be clear later from the examples given that loan items creating phonological problems in Bahasa Indonesia are, mainly, from Arabic (A), Dutch (D), English (E), Sanskrit (S), and Javanese (J).³ Items from other foreign (Fo) or regional (Ro) languages are very few.

The phonetic symbols used here are basically those of IPA. The exceptions are [j, š, y] which correspond in value with IPA [dʒ, ʧ, j] respectively.

2. DIVERSITIES DUE TO LOAN SOUNDS

The introduction of foreign words into Bahasa Indonesia (Malay) has brought with it some foreign sounds into the sound system of the language. These sounds, being uncommon, cause diversities in pronunciation, spelling, and to some degree, in the phonological description of Bahasa Indonesia. As I argued before (Lapoliwa 1981:25-28) four of these loan sounds ([f, z, š, x]) have to be posited as systematic phonemes in the underlying representation of loanwords in Bahasa Indonesia simply because the loan sounds are not always replaceable with their nearest corresponding native sounds. Examples (1) below show that these sounds can be naturalised in some loanwords.

	Spelling	Source	Pronunciation	Gloss		
(1)	a. i.	pikir, fikir	(A)	[pikIr, fikIr]	'to think'	
		napas, nafas	(A)	[napas, nafas]	'breath'	
		hapal, hafal	(A)	[hapal, hafal]	'to memorise'	
	ii.	tarif, tarip	(D, E)	[tarIf, tarIp]	'tariff'	
		lafal, lapal	(A)	[lafal, lapal]	'pronunciation'	
		verbal, perbal	(D, E)	[fərbal, pərbal]	'verbal'	
	iii.	pabrik	(A)	[pabrIk, fabrIk]	'manufacture'	
		aktif	(D, E)	[aktIf, aktIp]	'active'	
		foto	(D, E)	[foto, poto]	'photo'	
	b.	i.	asas, azas	(A)	[asas, azas, ajas]	'principle'
			jasad, jazad	(A)	[jasat, jazat]	'body'
			tamsil, tamzil	(A)	[tamsIl, tamzil]	'message'
		ii.	zaman, jaman	(A)	[zaman, jaman, saman]	'era, period'
			jazirah, jasirah	(A)	[jazirah, jasirah]	'peninsula'
			azab, ajab	(A)	[azap, ajab, asap]	'torture'
iii.		zat	(A)	[zat, jat, sat]	'substance'	
		zenit	(E)	[zenit, jenit, senit]	'zenith'	
		zona	(E)	[zona, jona, sona]	'zone'	
c. i.	sah, syah	(A)	[sah, šah]	'legal'		
	laskar, lasykar	(S)	[laskar, laškar]	'soldier'		
	saraf, syaraf	(A)	[saraf, šaraf, sarap]	'nerve'		

	Spelling	Source	Pronunciation	Gloss
(1) c.	ii. masyarakat, masarakat asyik, asik	(S)	[maʃarakat, masarakat]	'society'
		(A)	[asIk, aʃIk, asIʔ]	'absorbed'
	iii. aras	(A)	[aras, araʃ]	'God's throne'
	syak	(A)	[ʃak, ʃak, saʔ]	'worried'
	syekh	(A)	[ʃEx, ʃEk, sEk]	'sheik'
d.	i. kabar, khabar	(A)	[kabar, xabar, habar]	'news'
	kurma, khurma	(A)	[kurma]	'grape'
	kewan, khewan	(A)	[hewan, xewan, kewan]	'animal'
	ii. akhir, akir	(A)	[axIr, akIr, ahIr]	'end'
	khusus, kusus	(A)	[xusUs, kusUs, husUs]	'special'
	khayal, hayal	(A)	[zayal, hayal]	'dream'
	iii. khas	(A)	[xas, kas]	'typical'
	agen	(D, E)	[agEn, ahEn, axEn]	'agent, agency'
	biologi	(D, E)	[biologi, biolohi, bioloxi]	'biology'

It can be seen from the above examples that /f/ can be replaced by /p/, /z/ by /j/ or /s/, /ʃ/ by /s/, and /x/ by /k/ or /g/ or /h/.⁴ The use of a certain native sound as a substitute for a certain foreign sound may not necessarily be the same from loan item to loan item and from speaker to speaker. Thus one speaker may use native sound substitutes with loan items like pikir [pikIr], asas [asas], sah [sah], kabar [kabar] etc., but use loan (foreign) sounds with items like tarif [tarIf], zaman [zaman], masyarakat [masarakat], akhir [axIr] etc. Another speaker, however, may tend to use loan sounds with the first series (i.e. fikir [fikIr], azas [azas], syah [sah], khabar [xabar]) but use native sound substitutes with the second series (i.e. tarip [tarIp], jaman [jaman], masarakat [masarakat], akir [akIr]).

The third groups in the above examples (1.a.iii-d.iii) clearly show that a certain letter may have two or more sound values in loanwords. Thus p and f can each be interpreted phonetically as [p, f], z as [z, j, s], sy and s as [ʃ, s] each, kh as [x, k, h], and g as [g, h, x]. Such assignment of phonetic values of the letters just mentioned is not limited to the third groups in the above examples. A language user may use the form pikir in writing but occasionally use [fikIr] in speaking.

When the spelling forms of the above loanwords (see Poerwadarminta 1976) are disregarded, the linguist may be inclined to posit the native sound substitutes as underlying elements for the respective loanwords. The 'foreign' variants can then be derived by means of phonological rules. However, when he encounters the following pairs, he cannot do away with the loan (foreign) sounds.

	Spelling	Source	Pronunciation	Gloss
(2) a.	i. vak	(D)	[fak]	'school subject'
	pak	(D, E)	[pak]	'package'
	ii. fakta	(E)	[fakta, pakta]	'fact'
	pakta	(E)	[pakta]	'pact'
b.	i. zeni	(Fo)	[zəni, səni]	'(army) engineers'
	seni		[səni]	'art'

	Spelling	Source	Pronunciation	Gloss
(2) b.	ii. azab	(A)	[azap, asap]	'torture'
	asap		[asap]	'smoke'
c.	i. syak	(A)	[ʃak, sak]	'worried'
	sak	(D, E)	[sak]	'sack'
ii.	syarat	(A)	[ʃarat, sarat]	'requirement'
	sarat		[sarat]	'overloaded'
d. i.	khas	(A)	[xas, kas]	'typical'
	kas	(D)	[kas]	'cashbox'
ii.	tarikh	(A)	[tarIk, tarIx]	'year'
	tarik		[tarIk, tarIʔ]	'to pull'

Each pair of the above items forms some sort of minimal pair. The existence of those minimal pairs (although they sometimes appear identical on the phonetic level) has been made the main basis for recognising those loan sounds as constituting different phonemes (cf. Halim 1974:175-177 (1981:135-137); Lapoliwa 1981:26-27). Furthermore, the labial fricative /f/ in the following items is hardly ever replaced with /p/.

(3)	fasilitas	(E, D)	[fasilitas]	'facilities'
	federasi	(E, D)	[federasi]	'federation'
	fakultas	(E, D)	[fakUltas]	'faculty'
	faktor	(E, D)	[faktOr]	'factor'
	fantasi	(E)	[fantasi]	'fantasy'
	filial	(E)	[filial]	'filial'
	veto	(E)	[feto, fEt0]	'veto'
	visa	(E)	[fisa]	'visa'
	vital	(E)	[fital]	'vital'
	variasi	(E)	[fariasi]	'variation'
	universitas	(E, D)	[unifersitas]	'university'

The next question to solve after the assignment of phoneme status to the above foreign (loan) sounds is the one related to the phonemic (underlying) representation of individual loanwords. With items which are clearly 'immune' to the naturalisation process like those given in examples (3), or whose naturalised forms may be in conflict with items that already exist like those given in example (2), the use of the loan sounds for the respective loan items on the phonological level of representation is quite appropriate. But with items like those in example (1), the choice between native sound substitutes and the corresponding loan sounds for the respective loan items on the phonological level of representation is not so simple when the question of standard and sub-standard or common and less common forms is taken into consideration. Forms (orthographic and phonetic) that come first in example (1) above are regarded as standard forms while those that come second or third are regarded as sub-standard forms.⁵ It can be seen from examples (1) above that some naturalised forms are considered as standard forms (e.g. i-items) but some are not (e.g. ii-items). In fact, many of the i-items in examples (1) above have been fully naturalised to the extent that the average speakers of Bahasa Indonesia do not realise that they are foreign words in origin – they pattern and behave like indigenous words. If a descriptive grammar should reflect the competence of the majority of the speakers of Bahasa Indonesia, the linguist can not treat the i-items in (1) and the like in the same way as he does the rest of the loan items existing in Bahasa Indonesia.

In brief, the phonological problems which arise from the introduction of loan sounds into Bahasa Indonesia through word borrowing lie in the choice of appropriate variants. A user of Bahasa Indonesia is required to recognise those variants and uses them appropriately both in writing and speaking, while a field linguist is required to recognise them and their appropriate use so that his grammar may rightly describe true usage among Bahasa Indonesia speakers (cf. Chomsky 1965:4; Chomsky and Halle 1968:3).

3. DIVERSITIES DUE TO THE VIOLATION OF MORPHEME WELL-FORMEDNESS CONDITIONS

Besides the diversities stemming from loan sounds discussed above, the introduction of certain loanwords into Bahasa Indonesia has also caused diversities in pronunciation and, to some degree, in spelling. Such diversities are mainly due to the fact that many of the loanwords violate certain conditions which systematically constrain the forms of the majority of Bahasa Indonesia words. The conditions which may cause diversities among speakers of Bahasa Indonesia can be grouped, for the purpose of discussion, under what Richard Stanley (1967) called positive conditions and morpheme structure conditions.

3.1. Diversities attributed to positive conditions

Positive conditions, following Stanley, express the canonical shapes, i.e. the general constraints on the sequence of elements (consonant and vowel), of underlying forms of the majority of morphemes in Bahasa Indonesia. If we disregard bound morphemes and interjection words, we may be able to capture the general patterns of the rest of the indigenous words in Bahasa Indonesia by the following formula (cf. Lapoliwa, 1981:41):

$$++ (C \ 0 \ V \ (C(C)) \ V \ (C \ R((C)V(C)R)) ++$$

where C = Consonants (including glides)

V = Vowels

++ = word boundary

R = reapply the elements between the parentheses when necessary

(...0...) = at least one item should be applied

The formula states among other things that:

- a) a word consists minimally of CV (e.g. di 'in, on, at', ke 'to') or VV (e.g. ia 'he, she')
- b) the maximum sequence of C is CC, and this CC sequence may occur only in medial position (e.g. ambil 'take', hasta 'lower arm')
- c) the maximum sequence of V is VV (e.g. dua 'he, she', baik 'good')
- d) a word may end either in C or V

Of the four positive conditions, only condition b) needs further consideration in connection with loanwords. Many loan items cause diversities among speakers of Bahasa Indonesia because they violate that very condition. Example (4) below illustrates some diversities in the pronunciation of certain loanwords. The variants that come first in the series are considered as standard forms in that they satisfy the rules of the current spelling system and the guidelines for the coining of terminology.⁶

	Spelling	Source	Pronunciation	Gloss
(4) a.	linik	(D, E)	[klinIk, kəlinIk]	'clinic'
	tradisi	(D, E)	[tradisi, təradiSi]	'tradition'
	prihatin	(J)	[prihatIn, pərihatIn]	'to be concerned'
	struktur	(D, E)	[strUktur, sətrUktUr, stərUktUr, sUtrUktUr]	'structure'
b.	industri	(D? E)	[indUstri, IndUstəri]	'industry'
	instalasi	(D, E)	[Instalasi, Intalasi]	'installation'
	instruksi	(D, E)	[InstrUksi, InstərUksi, IntrUksi, IntərUksi, InsturUksi, InturUksi]	'instruction'
	konstruksi	(D, E)	[konstrUksi, konstərUksi, kOntrUksi, konstərUksi, kOnsturUksi, kOnturUksi]	'construction'
c.	teks	(D, E)	[tEks, tEk]	'text'
	ons	(D, E)	[Ons, On]	'ounce'
	korps	(D, E)	[kOrps, kOrp, korəp(s), korOp(s)]	'corps'
	helm	(D, E)	[hElm, helEm, heləm]	'helm'
	golf	(E)	[gOlF, goləf, golOf, goləp, golOp]	'golf'
	bank	(D, E)	[baŋk, baŋ]	'bank'
	tank	(D, E)	[taŋk, taŋ]	'tank'
	sport	(D, E)	[sport, spor]	'sport'
	tart	(D, E)	[tart, tar]	'tart'

The above examples show that unpermitted sequences of consonants tend to be broken up by vowel insertion and/or consonant deletion processes. In most cases, a schwa is used to break up unpermitted sequences of consonants that are found in loanwords. This may be due, partly, to the fact that the schwa in indigenous words is often deleted in rapid speech, in some of the above examples, however, we see that an unpermitted sequence of consonants can be broken up by a certain vowel which is very much similar to the one immediately following it (e.g. [U] in *instruksi*, *konstruksi*) or preceding it (e.g. [O] in *korps*, *golf*, and [E] in *helm*). The insertion of [U] in the phonetic realisations of the words *instruksi*, *konstruksi*, *struktur*, [O] in the words *golf*, *korps*, and [E] in the word *helm* is clearly a case of assimilation, or, more specifically, vowel harmony. The fact that the schwa ([ə]) also occurs in those items makes the task of the linguist harder. The question he needs to answer is concerned with the order of rules applied to derive the phonetic forms [IntUrUksi], [kOntUrUksi], [sUtrUktUr], [korOp(s)], [golOf, golOp], [helEm] etc. Unless one has convincing evidence that these forms are used by a certain group of speakers only, he cannot discount the order of rules deriving the phonetic forms. Since [u, o] in Malay words are never realised as [ə] on the phonetic level, it is more logical then to assume that in the breaking up of unpermitted consonant sequences schwa insertion applies first, then the inserted schwa is harmonised with the adjacent vowel. The rules for the vowel insertion can be formulated as follows.

R1 - *schwa insertion*

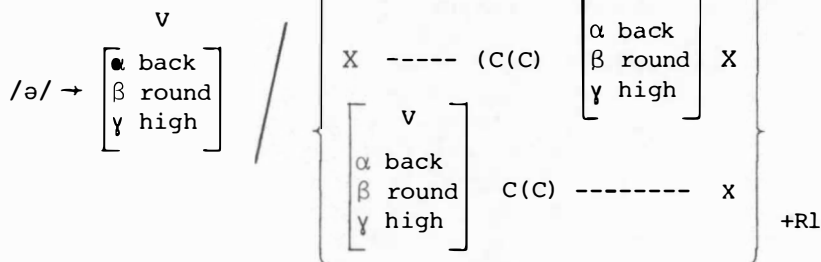
∅ ----> [ə] / ++ XC --- C(CV(C))++

where C = any consonant

V = any vowel

X = any segment sequence or zero

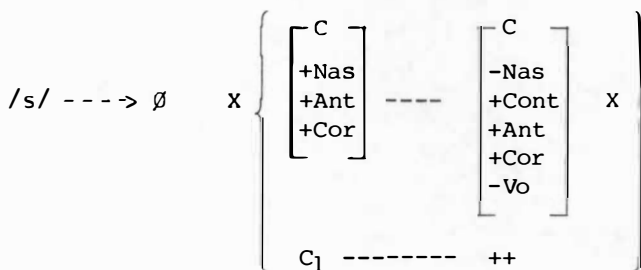
R2 - vowel harmony



where C = any consonant
 V = any vowel
 X = any sequence of segments

The prevention of unpermitted patterns (consonant sequences) by means of deletion does not seem to be very general. The conclusions that can be drawn from the above examples are (1) consonant sequences in word-initial position do not undergo deletion, (2) the consonant [s] in the sequence NASAL + /s/ + STOP in medial position tends to be deleted when both the nasal and the stop are homorganic with /s/ (+ANTERIOR, +CORONAL) (this condition is necessary because /s/ in *inspirasi* 'inspiration', *inspeksi* etc. is never deleted), (3) the fricative /s/ and stops preceded by sonorants in final position tend to be dropped.⁷ The rule for the /s/ deletion can be formulated as follows.

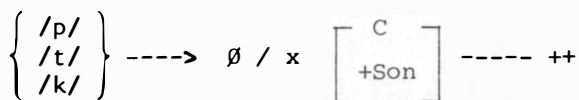
R3 - s-deletion



where C₁ = any consonant other than liquid
 X = any segment sequence or zero

The rule for the deletion of stops in final position can be formulated as follows.

R4 - stop deletion



where C = sonorant consonant other than lateral
 X = any sequence of segments

So far spelling forms shown in example (4) above look very simple. Users of Bahasa Indonesia are in fact as divergent in spelling as they are in pronunciation. The orthographic forms given above happen to be the ones that appear in the latest edition of *Kamus Bahasa Indonesia* (Poerwadarminta 1976), the most reliable dictionary available at present. Forms like *kelinik* or *clinic* besides *klinik*, *tek* or *text* besides *teks*, *perihatin* besides *prihatin*, *intalasi* besides *instalasi*, etc. are not rare in the general use of the loan items. Those foreign forms are,

of course, used only by those who happen to be familiar with the source language(s), while those reduced forms are usually used by less educated people. The forms with the inserted e do not tell the background of their users. The reasons for their use might be due to the fact that the user, either consciously or unconsciously, treats them as native words, or that he had been acquainted with those forms or similar forms before the promulgation of the present spelling system of Bahasa Indonesia. Indeed, there had been attempts made to naturalise 'ill-formed' loanwords. The existence of a number of doublets which are still found in the latest edition of *Kamus umum Bahasa Indonesia* shows that such naturalisation efforts had been made before. The doublets presented in example (5) below are all taken from Poerwadarminta (1976).

Spelling	Source	Pronunciation	Gloss
(5) klise, kelise	(D)	[klise, kəlise]	'film negative'
film, pilem	(D, E)	[fIlm, piləm, pilEm]	'film'
praktek, peraktek	(D, E)	[praktEk, pəraktEk]	'practice'
prangko, perangko	(D)	[pranko, pəranko]	'stamp'
status, setatus	(D, E)	[statUs, sətətUs]	'status'
stasiun, setasiun	(D, E)	[stasiUn, sətasiUn]	'station'
amplop, empelop	(D, E)	[amplop, ampəlOp]	'envelope'
staf, setap	(D, E)	[staf, sətap]	'staff'
prajurit, perajurit	(S)	[prajurIt, pərajurIt]	'private'
statis, setatis	(S)	[statIs, sətətIs]	'static'

The spelling (and pronunciation) forms that come first in the above examples are regarded as standard forms because they are in accord with the rules of the current spelling system and the guidelines for the coining of terminology mentioned above. The other variants – the naturalised ones which were regarded as standard forms at one time – are considered substandard. This phenomenon in which naturalised forms of loanwords serve no longer as standard may be referred to as a *denaturalisation* process. The denaturalisation of loanwords in Bahasa Indonesia is not without good reasons, but the problems it entails are not less serious. In addition to learning these new foreign forms, the users have to get rid of the unwanted, but nonetheless more natural, habits of speaking and writing.

3.2. Diversities attributed to morpheme structure conditions

Morpheme structure conditions to be discussed in this section will be limited to sequential conditions that are much violated by the introduction of foreign words into Bahasa Indonesia and as the result of such violation of conditions speakers of Bahasa Indonesia diverge in using the loan items. Sequential conditions are systematic constraints that a language imposes on the distribution of sounds.

Diversities among speakers of Bahasa Indonesia because of the violation of conditions on the distribution of consonants have been discussed in section 2 and subsection 3.1. above. The remaining discussion in this subsection will be concerned with diversities among Bahasa Indonesia speakers attributable to the violation of constraints on the sequential distribution of vowels in Bahasa Indonesia resulted from borrowing.

Vowels that may occur in word-final syllables in Malay (Bahasa Indonesia) are limited to /i, e, a, o, u/, but their distribution is very much conditioned by the vowel in the second-last syllable. The interdependence of vowels in the

last two syllables of the majority of Bahasa Indonesia words can be seen in the following table.

Penult	Ultima	Examples		
/i/	/i, u, a/	ini	/ini/	'this'
		ibu	/ibu/	'mother'
		tidak	/tidak/	'no, not'
/u/	/i, u, a/	tuli	/tuli/	'deaf'
		pukul	/pukul/	'to beat'
		dua	/dua/	'two'
/ə/	/i, u, a/	beri	/bəri/	'to give'
		betul	/betul/	'right'
		besar	/besar/	'big'
/a/	/i, u, a/	mati	/mati/	'dead, die'
		baru	/baru/	'new'
		makan	/makan/	'to eat'
/e/	/e, o, a/	pendek	/pendek/	'short'
		besok	/besok/	'tomorrow'
		merah	/merah/	'red'
/o/	/e, o, a/	boleh	/boleh/	'may'
		bodoh	/bodoh/	'stupid'
		ombak	/ombak/	'wave'

We see from the above table that [ə] never occurs in the word-final syllable. For this reason quite a number of loan sounds containing [ə] in the last syllable were phonetically realised as [e] or [a] as illustrated by examples (6) and (7) below respectively.

Spelling	Source	Pronunciation	Gloss
(6) sistem	(E)	[sɪstəm, sɪstɛm]	'system'
tablet	(D, E)	[tablət, tablɛt]	'tablet'
filter	(E)	[fɪltər, fɪltɛr]	'filter'
ganteng	(J)	[gantəŋ, gantɛŋ]	'handsome'
bandel	(J)	[bandəl, bandɛl]	'disobedient'
pilek	(J)	[pilək, pilɛk]	'flu'

The vowel /ə/ in the last syllable of the above items is never to my knowledge realised as [a]. Items in the following examples show that /ə/ may be realised as [e] or [a] besides [ə].

Spelling	Source	Pronunciation	Gloss
(7) akte	(D)	[aktə, akte, akta]	'document, certificate'
tante	(D)	[tantə, tante, tanta]	'aunt'
zone, zona	(E)	[zonə, zone, zona]	'zone'
ritme, ritma	(D? E)	[rɪtmə, rɪtme, rɪtma]	'rhythm'
ruwet	(J)	[ruwət, ruwɛt, ruwat]	'complicated'
serem, seram	(J)	[sərəm, serɛm, sərəm]	'frightening'

The rules accounting for the naturalisation of final /ə/ can be formulated as follows.

R5 - naturalisation of final /ə/

$$/ə/ \text{ ----> } \left\{ \begin{array}{l} [e] \\ [a] \end{array} \right\} / X \text{ ---- (C) ++}$$

where C = any consonant

X = any sequence of segments

The problem concerning those variants is in selecting or recognising standard forms. The tendency to select the more 'authentic' forms may not always be true because some forms like *seram* [səram], *zona* [zona], though less authentic, are now regarded as standard either by virtue of linguistic rules or because of other factors.

It can also be seen in the table above that when /ə/ occurs in the penult, the vowel of the ultima must be either a high vowel or a low vowel. This condition has made, to some degree, speakers of Bahasa Indonesia divergent in the pronunciation of a couple of loanwords. The pronunciation diversities may be due, partly, to the fact that the letter e represents the vowels [e] and [ə].

	Spelling	Source	Pronunciation	Gloss
(8) a.	beton	(Fo)	[bət0n, bet0n]	'concrete'
	germo	(J)	[gərm0, gErmo]	'hunter'
	jebol	(J)	[jəb0l, jeb0l]	'destruct'
	pergok	(J)	[pərg0k, pErg0k, pErg0?]	'to meet (somebody) incidentally'
	sedot	(J)	[səd0t, sed0t]	'to suck'
b.	semen	(D, E)	[səmEn, semEn]	'cement'
	parlemen	(D, E)	[pərləmEn, parlemEn]	'parliament'
	persen	(D, E)	[pərsEn, pErsEn]	'percent'
	permen	(D, E)	[pərmEn, pErmEn]	'peppermint'
	elemen	(D, E)	[eləmEn, elemEn]	'element'

From the above examples we see that the penultimate e (/ə/) is often realised phonetically as [e] when the last syllable contains a non-central mid vowel. This phonetic realisation can be expressed by the following rule.

R6 - fronting of /ə/

$$/ə/ \text{ ----> } [ə] / X \text{ --- C(C) } \left[\begin{array}{l} V \\ -\text{high} \\ -\text{low} \\ \alpha\text{back} \\ \alpha\text{round} \end{array} \right] \text{ (C) ++}$$

where C = any consonant

X = any segment sequence

There is still one condition related to the distribution of /e/, or rather the interpretation of the letter e, which may be interesting to note. As shown in the above table, /e/ in a penultimate syllable in Bahasa Indonesia (Malay) is usually followed either by a non-central mid vowel or by /a/. This is interesting because the modification made to the word may not be limited only to the penultimate vowel but also to the last vowel. The number of loan items behaving this way, however, is very limited. The following are some that I manage to record.

Spelling	Source	Pronunciation	Gloss
(9) semi	(E)	[semi, səmi]	'semi'
pening, peneng	(D)	[peniŋ, pəniŋ, penɛŋ]	'a metal badge (for dog)'
pensil	(E)	[pɛnsɪl, pənsɪl, pɛnsɛl]	'pencil'
seri	(D, E)	[seri, səri]	'series'
But: bendi	(Ro)	[bɛndi]	'(horse) cart'

The above examples show that there is no observed regularity concerning the modification made to loanwords containing /e/ in the penult and a high vowel in the ultima, except that such sequential order tends to be avoided on the phonetic level.

4. CONCLUSION

It may have been noticed from the examples given above that most of loanwords that cause diversities among speakers of Bahasa Indonesia when they use those items have come from English or Dutch, Arabic, and Javanese. The problem of diversities introduced by English/Dutch is not very much related to the four loan sounds - /f, z, ʒ, x/ - but rather to the structure of the loan items. The problem introduced by Arabic items is very much concerned with the loan sounds, while items of Javanese origin, like those from English or Dutch, present problems related to word structures. As the result of the existence of a number of competing variants due either to uncommon sounds or to 'ill-formed' words, language learners and language users are likely to find difficulty with which of the competing variants to learn or use.

The significance of this study lies probably in the message it tries to put across to those involved in language planning and in comparative studies - well-formedness conditions of morphemes should not be overlooked. Individual sounds are easy to detect and, therefore, problems entailed can easily be overcome. The structures of patterns or their combinations are not so easy to detect, and therefore, the problems entailed by the violation of their constraints may not be very easy to overcome.

NOTES

1. I became interested in this topic when in the course of the writing of my M.Phil. thesis submitted to the University of London in 1977 (see Lapoliwa 1981) I came across quite a number of items, which turned out to be loanwords, that behaved differently from the major part of the corpus.
2. In collecting the data, I had to rely very much on my impressions of what I heard, and had not followed any sort of scientific procedures like interviewing, making recordings, or attesting the value of the utterances. In studying the data, I am fully aware of the fact that the existence of some of the variants may be attributed to regional or foreign language interference. The fact of the existence of diversities concerning loans among speakers of Bahasa Indonesia, however, remains true.
3. This also reflects that borrowings are mostly from those languages (cf. Lapoliwa 1981).
4. The naturalisation of /x/ as [g] is limited to words borrowed from Dutch.
5. The assignment of the status of those forms is linguistically vague. According to the guidelines for the Spelling System of Bahasa Indonesia, "Items that have been adopted into Bahasa Indonesia and commonly spelt like other Indonesian words need no alteration" (the translation is mine).
6. The rules set up, in those two booklets, concerning the adoption of foreign items are mainly about spelling.
7. It seems that /s/ in final position is not deleted when preceded by sonorants other than /n/ because /s/ in *mars* 'march', *wals* 'false (pitch)' is never deleted. The condition for the deletion of stops does not seem to apply to stops preceded by /l/ as observed in talk '(powder)'.

BIBLIOGRAPHY

CHOMSKY, Noam

1965 *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T. Press.

CHOMSKY, Noam and Morris HALLE

1968 *The sound pattern of English*. New York: Harper and Row.

HALIM, Amran

1974 *Intonation in relation to syntax in Bahasa Indonesia*. Jakarta: Lembaga Bahasa Nasional and Penerbit Jambatan. Republished, slightly amended, as

1981 *Intonation in relation to syntax in Indonesian*. PL, D-36.

LAPOLIWA, Hans

1981 *A generative approach to the phonology of Bahasa Indonesia*. PL, D-34.

PANITIA PENGEMBANGAN BAHASA INDONESIA

1977 *Pedoman umum ejaan Bahasa Indonesia yang disempurnakan*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

1976 Pedoman umum pembentukan istilah. *Bahasa dan Sastra* 1/4:1-40.

POERWADARMINTA, W.J.S.

1976 *Kamus umum Bahasa Indonesia*. Jakarta: Balai Pustaka.

STANLEY, Richard

1967 Redundancy rules in phonology. *Language* 43:393-436.

THE PHONOLOGY OF JAVANESE VOWELS

Colin Yallop

1. OUTLINE

There has been some disagreement about the description of standard Central Javanese vowels. This paper reviews the major complexities in the vowel system, namely

the relationship between *i* and *I* and between *u* and *U* (2.1.)

the relationship between *é* and *è* and between *ó* and *ò* (2.2.)

the lowering of *i* and *é* (to *è*) and of *u* and *ó* (to *ò*) in certain contexts (2.3.)

the relationship between unrounded *a* and rounded *â* (2.4.)

While it is possible to generalise – for example, *i*, *é*, *u*, *ó* are restricted to open syllables – the detailed facts are such that generalisations may be superficial and misleading. In fact observation of regional or non-standard dialects, which often turn out to be simplifications of the standard dialect, suggests that there is genuine tension or complexity in the vowel system of the standard language. It is therefore unwise to approach Javanese in a way that seeks to simplify or generalise too readily (as some models of phonology are inclined to). Instead it seems necessary to distinguish fairly clearly between phonological structure, morphological structure and the contrastive phonemic system.

Any language must in a sense integrate these different aspects (by, for example, allowing phonemes to be realised differently in different structural or morphological contexts). It may be argued that Javanese is rather unusual in its path of integration, particularly in the extent to which morphological considerations are allowed to constrain phonological structure and the realisation of phonemic contrasts: but this makes it all the more important to do descriptive justice to the ingredients of the solution.

(See Appendices 1 and 2 for sources and symbols.)

2. DISTRIBUTION AND ALTERNATIONS

2.1. The vowels *i*, *I*, *u* and *U*

2.1.1. General distribution

It is more or less true that *i* and *u* occur only in open syllables while *I* and *U* occur only in closed syllables, e.g.

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 299–319. *Pacific Linguistics*, C-75, 1982.

© Colin Yallop

pi.pi 'cheek'
 pi.pIh 'rag'
 tu.ku 'buy'
 pUr.na 'restored, complete'

Where affixation alters syllabic structure, alternations occur, e.g.

gá.rIng 'dry' gá.ri.ngâ 'even if dry'
 ká.sUr 'mattress' ká.su.ré 'the mattress'

2.1.2. Syllabification

The definition of open and closed syllables in Javanese is not entirely straightforward, however. Only a few word-medial consonant clusters, notably r plus consonant, divide so as to close the preceding syllable, e.g.

sIr.nâ 'vanished'
 prIk.sâ 'examine'
 pIr.sâ 'see, perceive'
 gUs.ti 'lord, master'
 kUr.si 'chair'

Thus a considerable number of consonant sequences seem to count as syllable initial, including not only homorganic stop plus nasal but also consonant plus r or l and even the cluster ngs, e.g.

ti.mbUl 'emerge'
 dhi.ngklIq 'stool'
 pi.ntěr 'clever'
 lu.mrah 'usual'
 pu.trâ 'child'
 mu.ngsUh 'enemy'

Some authors (e.g. Robson 1976) agree with the syllable divisions given here. But Uhlenbeck (1949:38) rejects the idea that vowel variants can be explained purely in terms of syllabification. He suggests, for instance, that the syllable boundary does fall between a nasal and following stop, i.e. that pintěr is pin.těr. In this case it is obviously necessary to state that the 'open-syllable' variants i and u occur in open syllables and before certain consonant sequences. But in fact Uhlenbeck indicates that it is possible to describe Javanese phonology without necessarily referring to syllables at all.

A further complication is that clusters of consonant plus y have an ambiguous status: compare wIryâ 'brave' and dibyâ 'powerful, invulnerable'. If syllabic structure is what conditions the distribution of i, I, u and U, then these two words must be wIr.yâ and di.byâ. (But Uhlenbeck, 1949:35, comments that words containing consonant plus y are mostly archaic or literary.)

2.1.3. Effects of -áké

The 'causative' suffix -áké (see Appendix 3 for suffixes) does not affect preceding I or U, despite the fact that it would appear to create an open syllable. Compare forms in -áké with forms in -é, -i, -ânâ:

Root	-áké	-é	-i	-ânâ	Root meaning
sálIn	sálInáké	sáliné	sálini	sálinânâ	'change'
cáwIs	cáwIsáké	cáwisé	cáwisi	cáwisânâ	'prepare, provide'
gětUn	gětUnáké	gětuné	gětuni	gětunânâ	'regret, sorrow'
tulUng	tulUngáké	tulungé	tulungi	tulungânâ	'help'

This apparent anomaly may be due to paradigmatic uniformity. The Javanese verbal system reveals three 'causative' suffixes and three 'locative' suffixes (Appendix 3). In general the allomorphs of these suffixes are such that a root-final closed syllable remains closed before a causative suffix but becomes open before a locative suffix. -áké ought to break this pattern (as a causative suffix creating an open syllable) but it actually conforms to the causative paradigm:

Root	Causative Forms			
sálIn	sálInáké	sálinnâ	sálinné	I throughout
gětUn	gětUnáké	gětUnnâ	gětUnné	U throughout
Locative Forms				
sálIn	sálini	sálinânâ	sálinané	I i in locative
gětUn	gětuni	gětunânâ	gětunané	U u in locative

Dudas (1976:175) adopts a similar 'paradigmatic' explanation but it should be noted that this is expressed as a 'statement' and not incorporated within the rule system itself.

The only other suffixes which begin with a vowel but nevertheless do not trigger adjustment of I and U are the Krama equivalent of -áké, namely -ákĕn, and the Krama equivalent of -é, namely -ipUn (cf. Soepomo, 1969:168). If paradigmatic pressure is at work, it is not surprising that -ákĕn parallels -áké; why -ipUn is also irregular is not clear, unless its Krama status is a factor (Uhlenbeck 1949:209, mentions that I and U *may* change before -ipUn but that the unchanged form is regarded as more elegant). In any event, it does not seem possible to offer any explanation in terms of the phonological shape of the exceptional suffixes; it seems misleading, for example, to suggest that I and U become i and u only before monosyllabic suffixes, since they do change before -ânâ and -áné, as illustrated above.

2.1.4. Other apparent exceptions

There are some special instances in which i and u can appear in closed syllables, in particular in loan words (e.g. pit 'bicycle' from Dutch fiets) and in certain stressed or intensified words (e.g. ciliq 'tiny', but ciliQ 'small', cf. Uhlenbeck 1949:31ff, 66ff., Horne 1974:xxvii). Such occurrences may be regarded as further evidence that the distribution of i, I, u and U is not determined solely by phonetic or phonological factors (such as whether a syllable is open or closed); rather, morphological and even lexical factors interact with what may seem at first sight to be a matter of simple phonetics.

2.2. The vowels é, è, ó and ò

2.2.1. General distribution

Again, the distribution of these vowels is partly but not entirely accounted for by distinguishing between open and closed syllables. To some extent é and ó occur in open syllables and è and ò in closed syllables, e.g.

ké.né	'here'
gó.lèq	'get'
ló.ró	'two'
wé.dòq	'female'
só.ré	'evening'

But è and ò do occur in open syllables if the vowel of the following syllable is ě or (word-final) i or u; è also occurs before è, and ò before ò. These conditions are not the same as for I and U, as exemplified by the following, where incorrect pronunciations are asterisked:

before ě	mèsĕm *mésĕm 'smile'	mbòtĕn *mbótĕn 'not'	*pĪrĕng pirĕng 'hear'	*sUkĕt sukĕt 'grass'
before i	dĕwi *dĕwi 'goddess'	kòri *kóri 'door'	*sĪji siji 'one'	*bUmi bumi 'earth'
before u	sĕwu *sĕwu 'thousand'	*wòlu *wólu 'eight'	*pĪtu pitu 'seven'	*lUcu lucu 'funny'
identical vowels	dĕrĕng *dérĕng 'not yet'	sòròt *sóròt 'beam, ray'	*pĪtĪq pitĪq 'chicken'	*jUpUq jupUq 'pick up'

2.2.2. Effects of affixation

The complementarity of é and è and of ó and ò holds only *within* a morpheme. Alternations do not arise even where affixation seems to invite them, e.g.

sáté-nĕn	= sáténĕn	'make saté, let saté be made'	(cf. mèsĕm above)
sótó-nĕn	= sótonĕn	'make soto, let soto be made'	(cf. mbòtĕn above)
wòng-é	= wòngé	'the man'	(cf. sóré above)
lèmĕq-é	= lèmĕqé	'the lining'	
sòròt-é	= sòròté	'the beam'	

(But note also 2.3.3. below.)

2.2.3. Overlap with I and U

Since é and ó are confined to open syllables, and I and U to closed syllables (subject to exceptions mentioned in 2.1.3.), it is possible to identify I and U as allophones of é and ó. This is in effect the analysis

followed by Horne (1963), Soepomo (1969) and Sumukti (1971), and represented by Column D in Appendix 2. The phonetic justification is clear enough, since I and U are often closer in auditory quality to é and ó than to i and u. In an analysis which claims to be psychologically realistic, the question is how far native speakers' sensitivity to this auditory similarity is outweighed by their awareness of alternations (e.g. I can become i but not é; i.e. salIn-i is pronounced sálini not *sálèni or *sáléni). Authors who adopt what is here called Analysis D (Appendix 2) are of course obliged to introduce a process of raising, whereby, for example, salén + i → salini.

2.2.4. Contrast between é and è, and between ó and ò

Even if we do *not* take I and U to be allophones of é and ó, there is still a difficulty in grouping é and è together. We may in fact consider é and è as contrasting vowels — provided that we ignore morphological structure. Because of the facts mentioned in 2.2.2., it is possible to find both ..CéCé and ..CèCé in Javanese. But the second of these can occur only where the final é is the definite suffix (as in lèmèqé 'the lining').

A similar point can be made about ó and ò, where, for instance, ..CóCé occurs where there is no morpheme boundary before the é (e.g. sóré 'evening') but ..CòCé where -é is a suffix (e.g. sòròté 'the beam').

Minimal pairs are not very common. Potentially at least, there is a contrast between such pairs as bágéné (bágé-né) and bágèné (bágèn-é), both based on bágé 'share, distribute'. A minimal pair for ó and ò is kèbó-né 'the buffalo' and kèbòn-é 'the garden'. (But since â coincides with ò in the standard language, there are further examples of contrast, as in 2.4.2. below.)

2.3. Vowel lowering

2.3.1. General statement

Before the suffix -n and causative and locative suffixes (see Appendix 3), root-final vowels are lowered as follows

i and é become è
u and ó become ò

Examples:

Root	-n	Causative -qáké	Locative -ni	Root Meaning
isi	isèn	isèqáké	isèni	'contain'
bágé	bágèn	bágèqáké	bágèni	'distribute'
těmu	těmòn	těmòqáké	těmòni	'find, meet'
páro	páròn	pàròqáké	pàròni	'half'

The other causative and locative suffixes cause identical lowering, e.g.

isèni, isènanà, isènané, etc.

2.3.2. Morphological conditioning

Although partly constrained by pronounceability, lowering is not just a response to phonetic requirements. For example, the root-final vowel of *isi-n* must undergo some modification, since *i* ought not to appear in a closed syllable; but *isIn* is as readily pronounceable as *isèn* (and *isIn* does in fact exist as the word '*shy, embarrassed*'). Moreover, there are suffixes of comparable shape, such as *-nèn*, *-né* and *-mu*, which do not trigger lowering, e.g.

Root	With Lowering			Without Lowering		
	-n	-ni	-náné	-nèn	-né	-mu
<i>isi</i>	<i>isèn</i>	<i>isèni</i>	<i>isènané</i>	<i>isinèn</i>	<i>isiné</i>	<i>isimu</i>
<i>bágé</i>	<i>bágèn</i>	<i>bágèni</i>	<i>bágènané</i>	<i>bágénèn</i>	<i>bágéné</i>	<i>bágému</i>

Note in particular that some of the suffixes that do not require lowering actually create sequences that ought in a sense to be unpronounceable, such as *é* before *è*, or *é* before final *u* (cf. 2.2. above). Thus

- (i) lowering applies before a set of suffixes which cannot be defined in purely phonetic or phonological terms;
- (ii) lowering does more than make the minimal phonetic adjustment necessary to achieve an acceptable Javanese pronunciation;
- (iii) lowering does not apply in some contexts, where the resulting form appears to violate normal rules of pronunciation.

2.3.3. Interaction with constraints on *é*, *è*, *ó* and *ò*

Where a root-final *é* or *ó* is preceded by itself, both vowels must be lowered before the relevant suffixes, e.g.

<i>léndhé-ni</i>	=	<i>lèndhèni</i>	'lean on'
<i>ngéné-qáké</i>	=	<i>ngènéqáké</i>	'do/make in this way'
<i>bódhó-n</i>	=	<i>bòdhòn</i>	'ignorantly'
<i>ngónó-qáké</i>	=	<i>ngònòqáké</i>	'do/make in that way'

This is obviously a straightforward consequence of the constraint against having *é-è* or *ó-ò* within a morpheme (2.2.1.). If lowering applies to the second of two *é*'s or *ó*'s then the preceding vowel (which is within the same morpheme) must also be lowered.

2.4. The vowels *á*, *à* and *â*

2.4.1. General distribution

á and *à* are complementary, in open and closed syllables, e.g.

<i>gá.rIng</i>	'dry'
<i>ó.màh</i>	'house'
<i>má.ngàn</i>	'eat'
<i>mé.já.né</i>	'the table'

As with *é*, *è*, *ó* and *ò* (2.2.2.), the distribution of *á* and *à* is immune to affixation, e.g.

<i>ómàh</i>	'house'	<i>ómàhé</i>	'the house'
-------------	---------	--------------	-------------

But, in the standard language, *á* has undergone backing and rounding to *â* if it is word final or if it stands in an open syllable preceding *â*, creating alternations such as

mé.jâ	'table'	mé.jâ.né	'the table'
wâ.câ	'read'	wá.cá.něn	'read it, let it be read'
ně.gâ.râ	'country'	ně.gâ.râ.ku	'my country'

2.4.2. Contrastive status of *â*

Since the vowel *â* (in open syllables) coincides in quality with the *ò* vowel (mainly in closed syllables, 2.2.), a strictly phonemic analysis must take the historically enlarged phoneme *ò* to be in contrast with *ó*, e.g.

lóró	'two'	lârâ (= lòrò)	'sick'
puló	'island'	kulâ (= kulò)	'I'
ngónó	'thus'	ngânâ (= ngònò)	'thus'

(these two forms differ deictically)

póló	'head, brain'
pólâ (= pólò)	'pattern'
pâlâ (= pòlò)	'nutmeg'

2.4.3. Syllabification

If one insists that non-final *â* occurs only in an *open* syllable, any rule or predictive statement must take account of the way in which consonant clusters are allocated to syllables (as with *i*, *I*, *u* and *U*, 2.1.2.), e.g.

tâ.mpâ	'receive'
kâ.ncâ	'friend'
sâ.srâ	'thousand'
bâ.ngsâ	'nation'

BUT	tân.pâ	'without'
	wâr.nâ	'colour'
	sàs.trâ	'literature'
	jâl.mâ	'human being'

Uhlenbeck's remarks apropos *i* and *u* are relevant, however (2.1.2.). Soepomo (1969:167) also does not explain penultimate *â* entirely in terms of open syllables: he says that penultimate *â* occurs in an open syllable or in a syllable ending in a nasal. On his interpretation 'friend' is therefore *kân.câ*, 'nation' is *bâng.sâ*. On the other hand, Sumukti (1971) has syllable divisions such as *tâm.pâ* but *tâ.mbâ*.

2.4.4. Words of more than two syllables

Viewed as a rule, rounding of the vowel *a* is not iterative, i.e. it does not proceed indefinitely leftwards. Hence *câ.râ.kâ* 'messenger', not **câ.râ.kâ*; *tâ.ri.mâ* 'accept', not **tâ.ri.mâ*, etc.; although Uhlenbeck (1949:39) does give *nâyâkâ* as an alternative to the more usual *nâyâkâ* 'councillor, State official'.

But rounding does apply to both elements of a compound, e.g.

kâ.lâ.mâ.nggâ (i.e. kâlâ#mânggâ) 'spider (variety of)'
ku.lâ.wâr.gâ (i.e. kulâ#wârgâ) 'family'

It is doubtful whether all words of this type are felt to be compounds by native speakers; but root morphemes of three or more syllables are in any case rather rare in Javanese. There would be no principial objection, for example, to saying that any Javanese root of four syllables is treated as two bisyllabic constituents, whether or not native speakers are aware of its compound origin (cf. Kiliaan, 1919:41, Uhlenbeck, 1949:203). It is significant that if a root is reduplicated and carries a suffix, there is variation in pronunciation, as illustrated by Soepomo (1969:167) with the two possibilities bu.tâ.bu.tá.né and bu.tá.bu.tá.né. The first alternative may be said to reflect the tendency to treat each half of the word independently, while the second overrides this tendency in recognition of the reduplicated character of the word.

2.4.5. Suffixes

Suffixes of appropriate form may contain â, i.e. -a is -â, -ana is -ânâ. But -â, which can follow a root-final vowel, does not block rounding of preceding a's, e.g.

bisâ	'can, be able'	bisââ	'even if possible'
pirâ	'how many'	pirââ	'however many'
lârâ	'sick'	lârââ	'even if sick'

(Compare 2.4.1. for the blocking effect of other suffixes such as -né, -něn, etc.)

-â is the only Javanese suffix which allows a vowel to follow a root-final vowel; other affixes have allomorphs with an intervening consonant (Appendix 3). Berg (1937:111) notes a further oddity, namely that â is raised to ó before the suffix -a, i.e. lârââ is pronounced lóróâ; Uhlenbeck (1949:208) mentions the same phenomenon with the example pirówâ for pirââ. Subsequent descriptions of Javanese do not take this up, and the present generation seem to regard the pronunciation with ó as archaic.

2.4.6. Summary

The occurrence of â is the result of a historical change whose phonetic motivation is, to say the least, obscure in the modern language. Note that where penultimate â precedes â, the rounding agreement is not a simple matter of phonetic vowel harmony, for there are actually two conditions under which an unrounded a can precede a low back rounded vowel:

- (i) where the second vowel is ò rather than final â, e.g. tá.wòn 'bee', pá.ròn 'half and half' and
- (ii) where à is suffixed to a root ending in ..àC (in which case -â apparently fails to create an open syllable) e.g.

kápàn	'when'	kápánâ	'no matter when'
nákàl	'naughty'	nákàlà	'even if naughty'

Moreover, if one pursues the idea of expressing rounding as a rule, the

structural conditions are quite difficult to state: within a single morpheme a sequence of three *â*'s is not possible (note **cârâkâ* in 2.4.4.) but both 'compounds' and affixation of *-â* do allow such a sequence (*kâlâmânggâ*, *lârââ*). Thus rounding might be expressed as a process which applies only to a sequence ...*(a.C)a#* within a morpheme, subject to the conditions that 'compounds' may contain two such sequences, both eligible to undergo the rule, and that the suffix *-â*, which itself undergoes the rule, does not prevent a preceding sequence from undergoing the rule. In fact Dudas resorts to two rules, one to account for rounding of final *-â*, the other to 'harmonise' a preceding *a*. Neither rule really deals with rounding before the suffix *-â*.

In this connection, it should be noted that there is no obvious formal device for marking the peculiarity of the suffix *-â*. It is not plausible, for instance, to suggest that *-â* is necessarily preceded by a # boundary (i.e. that *-â* is by nature less closely bound to a root than other suffixes). While this ploy might 'explain' why the suffix does not block rounding of preceding *a*'s, it overlooks the fact that *-â* is like other suffixes in creating open syllables in respect of allophones of *i* and *u*, e.g. *gârIng*, *gâringâ* (2.1.1.).

3. REGIONAL AND NON-STANDARD VARIATION

There are versions of Javanese pronunciation, some already touched on above, which eliminate various complexities. These pronunciations can in general be characterised either as regional or as non-standard.

In East Java there are speakers who use *I* and *U* in all word-medial positions and restrict *i* and *u* to word-final position, e.g.

tImbUl 'emerge'
pIndó 'twice'
kUpIng 'ear'
 (cf. 2.1.1., 2.1.2.)

This pronunciation represents a simplification in a number of respects:

the ambivalent status of *I* and *U* is resolved, for in this dialect they are clearly in contrast with *é* and *ó* and must therefore be variants of *i* and *u*; secondly, the distribution of *I* and *U* is no longer constrained by syllabic organisation or internal word structure (cf. 2.1.2.); and forms in *-âké* such as *sáInáké* are no longer exceptional (cf. 2.1.3.).

Even speakers who do not follow this East Javanese pronunciation show some inclination to simplify the rules of syllabification. The cluster *ngs*, for example, which seems an unlikely sequence to begin a syllable (2.1.2., 2.4.3.), is in fact often eliminated: for the words *mu.ngsUh*, *mâ.ngsâ*, *má.nu.ngsâ* and *sò.ngsòng*, for instance, there are variants *mu.sUh*, *mâ.sâ*, *má.nu.sâ*, and *sò.sòng*.

A second important area of variation concerns *â*. In some western areas the historical change of *á* to *â* has not applied, while in eastern areas it has generally been carried through even where a suffix follows, e.g.

Western	Central	Eastern	Meaning
<i>méjá</i>	<i>méjâ</i>	<i>méjâ</i>	'table'
<i>méjáné</i>	<i>méjáné</i>	<i>méjáné</i>	'the table'
<i>négárá</i>	<i>négârâ</i>	<i>négârâ</i>	'country'
<i>négáráku</i>	<i>négârâku</i>	<i>négârâku</i>	'my country'

Even Central Javanese speakers tolerate increasing exceptions to the distribution of *á* and *â*. A long-standing exception is *órá* 'not', which is listed as an irregularity in most grammars and dictionaries. Uhlenbeck also mentions *bóyá* (noting that negatives are often irregular) and some other exceptions, albeit archaic words (1949:31ff.). Furthermore, words in -a taken over into modern Javanese do not undergo rounding. Words with an Indonesian (i.e. national or official) flavour, such as the name of the country itself or the word *sěpédhá* 'bicycle' (versus colloquial *pit*) are pronounced with -a. The evident willingness of Javanese speakers to write *â* as *o* rather than *a* also suggests that the historical connection with *a* is receding and that the *a/o* alternation is felt to be a more or less arbitrary morphological feature rather than as an automatic or productive phonological process.

There is a real sense in which the standard dialect is intermediate between two simpler dialects – and it seems fair to say that this provides a motive for maintaining the alternation as a signal of the distinctiveness of Central Javanese. Javanese speakers who move to Central Java from elsewhere seem to feel some social pressure to acquire the alternation. Distinctive compromises of this sort seem to occur elsewhere, sometimes as a virtually artificial standard, as in the case of the German suffix -ig, where the standard pronunciation (with e.g. *König* pronounced as *Könich*, but *Königreich* and *königlich* as *Könikreich* and *köniklich*) presupposes the existence of some dialects with consistent -ich and others with consistent -ik. But a similar phenomenon seems to be possible even when there is no pedagogical interference. In a group of Australian Aboriginal dialects known as Pitjantjatjara or Western Desert, there is a dialect in which dental sounds such as *t̪* and palatal sounds such as *tʲ* are in a rather complicated complementary distribution: roughly, the dental stands before *a* and *u* unless preceded by a non-dental consonant, while the palatal stands before *i*, and also before *a* or *u* if preceded by a non-dental consonant (Glass and Hackett 1970: 109f.). This somewhat puzzling distribution of allophones becomes less mysterious when one realises, firstly, that there is some evidence that dental and palatal consonants in Australian languages may have developed out of a single laminally articulated series, with lamino-palatal allophones appearing before *i* and lamino-dental allophones before other vowels (Dixon 1970), and, secondly, that in the particular case of Western Desert there are other dialects in which all laminals have become dentals and yet others in which all laminals have become palatals. Thus the dialect in which a dental and a palatal consonant are allophones of one phoneme is actually intermediate between two simpler or 'levelled' versions of the language.

It is of course proper that linguists should seek to explain linguistic phenomena and to look for regularity behind apparent complexity. Often a language does prove to have an underlying symmetry or regularity that is not apparent on a cursory examination. But the evidence of variation in Javanese is significant, since it suggests that native speakers themselves find the standard dialect complex, both in the sense that regional and non-standard dialects represent simplifications and in the sense that speakers of a distinctive or prestige dialect have reason to resist simplification.

4. ASPECTS OF PHONOLOGY

I am not of course claiming that standard or prestige dialects are necessarily complex, and non-standard dialects necessarily simple. In fact it seems more realistic to suppose that all phonological systems are in a state of tension or represent a particular resolution of a state of tension (cf. recent attempts by natural phonologists to explain phonological systems as the result of competition between distinctiveness and ease of articulation, e.g. Stampe 1979:69f). This makes it important to do justice to the ingredients of such tension. Possibly all that can be said about prestige dialects, especially those that are taught formally, is that they are more prone to delicate or even awkward compromises.

Even apart from this, the details reviewed in section 2 above should make us cautious about offering too simple a picture of Javanese. There is, for example, little descriptive or pedagogical advantage in pairing off as many vowels as possible as tense and lax counterparts (say as *i*, *I*; *u*, *U*; *é*, *è*; etc.). The terms 'tense' and 'lax' are phonetically vague to begin with, but the relationship of these pairs is in any case not uniform. To say that tense vowels occur only in open syllables and lax only in closed syllables, for example, is an oversimplified generalisation, a rough approximation which sweeps considerable detail under the carpet (especially the details mentioned in 2.1.3., 2.1.4., 2.2.1. and 2.2.2. above). Of course, oversimplifications have their place, for example in an elementary language course, but they do not qualify as descriptive truth.

Rather than engaging in a reductionist exercise, let us distinguish in Javanese among (i) phonological structure (including word patterns and syllabification), (ii) morphological structure (i.e. morphemic composition) and (iii) the contrastive or phonemic system (i.e. the 'auditory network' as recognised and maintained by native speakers). Javanese reveals various resolutions of potential conflict among these aspects. For example, the normal pattern of word and syllable structure requires that CVCV and CVCVCV words such as *kéné* 'here' and *képâlâ* 'head' be understood as sequences of CV syllables. Morphological structure sometimes accommodates to this phonological structure, so that e.g. *sálIn* plus the suffix *-i* is analysed as *sá.li.ni*, with a consequent adjustment of *I* to *i*. But in other instances morphological structure resists 'normal' phonological patterning, as with e.g. *sálInáké* (2.1.3.) or *nákâlâ* (2.4.6.). It seems legitimate to say that morphological structure triumphs over phonological structure to quite a degree in Javanese. Native speakers seem, for instance, to regard *wòngé* 'the man' as *wòng.é*, despite the fact that a CVCV word consisting of one morpheme would be felt as CV.CV. Thus the morpheme boundary in *wòng-é* demands a syllable boundary at the same point (cf. Uhlenbeck, 1949:225ff.). In the standard dialect the tendency is for morphological structure to yield only where *I* and *U* alternate with *i* and *u* (with some exceptions such as *sálInáké*); regional and non-standard dialects tend to make the morphological structure even more dominant (with e.g. pronunciations such as *pitIq*, *pitIqé* '(the) chicken' instead of standard *pitIq*, *pitiqué*).

Tension between phonological structure and the contrastive system arises where phonemes are realised in different ways (or even neutralised) in different structural positions. Indecision about whether to write *â* as *a* or *ò* may be taken to reflect a conflict between a native speaker's awareness of structural patterns (words can end in *i*, *é*, *u*, *ó* or *â*, but not in *á*) and his auditory perception (*â* sounds the same as *ò*). As the structural pattern is broken, by more and more exceptions ending in *-á*, so the auditory system is likely to determine the spelling.

There are dangers here in promoting too singlemindedly some of the recognised models of phonological description. If, for instance, one insists that a structuralist description of phonemes and allophones provides the key to understanding the phonological system, then one runs the risk that phonological and morphological structure become secondary. The fact that not all types of morpheme or morphological structure show the same distribution of allophones (2.1.4., 2.2.2.) may too easily be seen as an odd intrusion upon the basic system. On the other hand, to insist upon a set of rules as the proper descriptive method is open to others dangers. It is of course possible to rewrite a statement of phonemes and allophones as a set of rules realising phonemes as allophones, and equally possible to include in the set rules that cater for morphophonemic adjustments. Indeed, wherever there is a resolution of conflict between two aspects of the language – say where morphemes have different allomorphs to accommodate to a regular phonological structure – it must in principle be possible to state this in terms of rules. (Uhlenbeck, writing well before generative phonology, does not object to the use of the term 'rule', even though his approach is very much one of presenting structural patterns, 1949:31.) The objection is not to the notion of rules but to the fashion of talking of 'a set of rules' in a way that obscures the different purposes which these rules serve. But since even generative phonologists seem increasingly willing to distinguish among 'phonetic rules', 'syllabification rules', 'morphophonemic rules', etc., the point perhaps needs no further elaboration (cf. Sommerstein 1977:205ff.).

This is not to say that one needs no model of description at all, nor that one should adopt a vague amalgam of models in the name of eclecticism. Rather it is important to reject the reductionist implications of some models and to base one's description on a clear recognition of the different aspects or dimensions of an integrated system. One may then allow for the fact that integration may proceed differently in different languages. In some languages, for example, each phoneme may be pronounced in much the same way in every structural environment, i.e. the auditory contrast system is relatively uniform with respect to phonological structure. In a language that has complex strings of morphemes which are modified to accord with a regular phonological structure, one may speak of a relatively high number of allomorphs or a relatively rich morphophonemics. In Javanese, on the other hand, with as Uhlenbeck puts it, rather few *voegverschijselen*, morphological patterns tend to resist both phonological structuring and the uniformity of the contrast system. The Javanese solution is perhaps unusual: one can point to contrasts which are contrasts only if morphology is ignored (é versus è, 2.2.4.); to rules which are blocked by the presence of a morpheme boundary (2.2.2., cf. Sommerstein's remarks (1977:145) on the rarity of this and on the difficulty of coping with it in a formal generative description); and to syllable patterns which deny the common tendency (e.g. wòngé, structured as CVC.V rather than CV.CV, cf. Bell and Hooper 1978, especially p.9 apropos the favourite status of the CV syllable).

This makes it all the more regrettable that the sort of description and pedagogy of Javanese outlined by Uhlenbeck (1949:13ff., 225ff.) and to some extent taken up by Sumukti (1971) has not been more honoured. Appendix 4 illustrates the type of tables or charts that were foreshadowed by Uhlenbeck and that might profitably be used to shed light on the Javanese phonological system.

For pedagogical purposes, one might have reservations about the use of such charts, on the grounds that they are too complicated. Nevertheless, one possibility is to choose one's initial examples of vowel contrasts fairly carefully and to ensure that new vocabulary is always introduced in sets of structurally

parallel items. Even if this were impracticable it might be worth devoting some time during a language course to a systematic survey of the phonology; this at least would be preferable to allowing students to stumble on in ignorance of structures and patterns which, though complex, can be revealed to them.

APPENDIX 1: USE OF JAVANESE MATERIAL AND ACKNOWLEDGEMENTS

In addition to the written sources listed among the references, the following individuals have been of great personal assistance in preparing this paper: Bintoro (Universitas Kristen Satya Wacana), Soekemi (IKIP Surabaya), Sumaryono (University of Sydney) and Urip Sutyono (Universitas Kristen Satya Wacana). I have tried to ensure that words and features of Javanese which I mention are substantiated by written sources and acceptable to native speakers, but I must make it clear that the four Javanese-speaking linguists named above are not to be blamed for any inaccuracies on my part.

Except at one or two points where there are clear morphological differences, I have ignored the sociolinguistic distinction between Ngoko and Krama (etc.) and have used words of any type for illustration. In general the vowel phonology (as opposed to, say, speech tempo or the lexical system) is identical for all speech levels.

A further simplification, which likewise ought not to affect discussion of the vowels, is that I have quoted verb forms without prefixes. Thus where the root *těmu* takes the 'locative' suffix *-(n)i* I have listed the resulting form as *těmòni* rather than as one of the forms more likely to occur, such as *němòni* or *ditěmòni*. This is merely to avoid the additional, and for present purposes irrelevant, complexity of dealing with prefixes and 'nasalisation' of root-initial consonants. It should also be noted that not all members of verb paradigms quoted in the paper are equally common. The reader is asked to accept that some forms are quoted to show how they would be pronounced in comparison with other forms, not necessarily because they are in common use.

APPENDIX 2: SYMBOLS

In this paper I write the Javanese consonants as in modern orthography, i.e. as p b t d th dh c j k g s h m n ny ng l r w y, where th and dh are postalveolar or retroflex plosives, c and j palatal plosives, ny a palatal nasal and ng a velar nasal; but I write k as q wherever it represents a glottal stop (in morpheme-final position). Where it is necessary to clarify structure, I use a full stop to indicate a syllable boundary, a hyphen for a morpheme boundary and # for a word boundary.

For the vowels I use twelve symbols to represent eleven phonetically distinct vowel qualities. The eleven vowels are usually grouped as six to eight contrasting or underlying phonemes:

A	B	C	D	Phonetic Description and I.P.A. Symbol
i			i	close front unrounded vowel [i]
I	i	i		lowered i; [ɪ] approaching [e]
é		é	é	half-close front unrounded vowel [e]
è	e	è	è	(half-)open front unrounded vowel; [ɛ] approaching [æ]
u			u	close back rounded vowel [u]
U	u	u		lowered u; [ɔ] approaching [o]
ó		ó	ó	half-close back rounded vowel [o]
ò	o	ò	ò	(half-)open back rounded vowel; [ɔ] approaching [ɒ]
â				used for the same vowel as ò where conservative spelling has a
á	a	a		open central unrounded vowel [a]
à		a	a	half-open central unrounded vowel [ɐ] or [ʌ]
ě	ě	ě	ě	half-close central unrounded vowel [ə] or [ɨ]

Column A gives the symbols used in this paper. Column B represents a six-vowel analysis, more or less as put forward by Uhlenbeck (1949) and Robson (1976 especially p.4). Dudas (1976) is a generative treatment in which the six vowels of column B are regarded as the underlying vowels. An eight-vowel analysis is rather more common, either in the manner of column C (as suggested by the transcription used in Uhlenbeck 1975) or in the manner of column D (as in Horne 1963:xvi-xviii or Soepomo 1969:167-168 or Sumukti 1971). Other analyses are possible, for example a seven-vowel analysis with *é* distinct from *è*, but *ó* and *ò* treated as allophones of *o* and *â* treated as an allophone of *a* (implied by Horne 1974:xi-xii). Of the works mentioned here, however, only Uhlenbeck (1949) is at all comprehensive: the others concentrate on matters other than phonology.

APPENDIX 3: PRINCIPAL SUFFIXES

Allomorphs			see note 1	see note 2	see note 3	meanings (see note 4)
after C	after V					
-àn	-n		yes	yes	yes	various
-áké	-qáké	see note 5	no	yes	yes	causative
-ná	-qná		(no)	yes	yes	causative subjunctive
-né	-qné		(no)	yes	yes	causative optative
-i	-ni		yes	yes	yes	locative
-áná	-náná		yes	yes	yes	locative subjunctive
-áne	-náne		yes	yes	yes	locative optative
-ěn	-něn	see note 5	yes	no	yes	passive imperative
-é	-né		yes	no	yes	definite
-ku			(no)	no	yes	'my'
-mu			(no)	no	yes	'your'
-â			yes	no	no	subjunctive

Notes

1. Do root-final -iC and -uC become -iC and -uC before the suffix?
2. Are root-final i and é lowered to è (and u and ó to ò) before the suffix?
3. Does the suffix block rounding of root-final a to â?
4. The labels given to meanings are, with slight simplification, taken from Horne 1974. Many of the labels are scarcely informative on their own and should be interpreted in the light of a fuller description of the suffixes (e.g. Berg 1937; Uhlenbeck 1956; Horne 1974).
5. In Krama usage, -(q)áké is replaced by -(q)ákén, and -(n)é by -(n)ipUn. Other suffixes either follow the substitution of -ipUn for -é (e.g. locative optative Ngoko -áne, Krama -ánipUn) or, in principle, are available in both Ngoko and Krama. (In practice some suffixes, such as -něn and -mu, would be avoided in Krama usage.)

APPENDIX 4: SAMPLE TABLES AND CHARTS FOR TEACHING JAVANESE

A. Word patterns

A.1. CV.CV words

First vowel: i, é/è, u, ó/ò, á/â, ě (6)

Second vowel: i, é, u, ó, â (5)

é and è in complementary distribution

ó and ò in complementary distribution

á and â in complementary distribution

Combinations:

i - i	i - é	i - u	i - ó	i - â
è - i	é - é	è - u	é - ó	é - â
u - i	u - é	u - u	u - ó	u - â
ò - i	ó - é	ò - u	ó - ó	ó - â
á - i	á - é	á - u	á - ó	â - â
ě - i	ě - é	ě - u	ě - ó	ě - â

Examples:

siji 'one'	piré 'avoid'	pitu 'seven'	biró 'bureau'	gilâ 'loathe'
dèwi 'goddess'	kéné 'here'	sèwu 'thousand'	réwó 'untidy'	sédâ 'dead'
bumi 'earth'	supé 'forget'	susu 'milk'	puló 'island'	kulâ 'I'
kòri 'door'	kówé 'you'	wolu 'eight'	lóró 'two'	tóyâ 'water'
máti 'die'	sáré 'sleep'	rátu 'king'	káró 'with'	lârâ 'sick'
wèdi 'afraid'	gédhé 'big'	lěbu 'enter'	jěró 'deep'	těkâ 'come'

A.2. CV.CVC words

First vowel: i, é/è, u, ó/ò, á, ě (6)

Second vowel: I, è, U, ò, à, ě (6)

é and è in complementary distribution

ó and ò in complementary distribution

C.2. Second vowel in CVCV

i	é	u	ó	â
puli	pulé	pulu	puló	
gili			giló	gilâ
etc.				

C.3. Root-final vowel before suffix -n (CVCVn)

è	ò	à	
lěgèn			lěgàn
lákèn	lákön		
etc.			

BIBLIOGRAPHY

BELL, A. and J.B. HOOPER, eds

1978 *Syllables and segments*. Amsterdam: North-Holland.

BERG, C.C.

1937 *Bijdrage tot de kennis der Javaansche werkwoordsvoormen. Bijdragen tot de Taal-, Land- en Volkenkunde* 95.

DIXON, R.M.W.

1970 Proto-Australian laminals. *Oceanic Linguistics* 9/2:79-103.

DUDAS, K.M.

1976 The phonology and morphology of modern Javanese. Ph.D. dissertation, University of Illinois.

GLASS, Anee and D. HACKETT

1970 *Pitjantjatjara grammar: a tagmemic view of the Ngaanyatjara (Warburton Ranges) dialect*. Canberra: Australian Institute of Aboriginal Studies.

HORNE, Elinor C.

1963 *Intermediate Javanese*. New Haven: Yale University Press.

1974 *Javanese-English dictionary*. New Haven: Yale University Press.

KILIAAN, H.N.

1919 *Javaansche Spraakkunst*. The Hague: Martinus Nijhoff.

POEDJOSOEDARMO, Soepomo

- 1969 Wordlist of Javanese non-Ngoko vocabularies. *Indonesia* 7:165-190.

ROBSON, S.O.

- 1976 A simple Javanese grammar. Mimeo. University of Auckland.

SOMMERSTEIN, Alan H.

- 1977 *Modern phonology*. London: Edward Arnold.

STAMPE, D.

- 1979 A dissertation on natural phonology. Mimeo. Indiana University Linguistics Club. Revised edn.

SUMUKTI, R.H.

- 1971 Javanese morphology and morphophonemics. Ph.D. dissertation, Cornell University.

UHLENBECK, E.M.

- 1949 *De structuur van het Javaanse morpheem*. *Verhandelingen van het Koninklijk Bataviaasch Genootschap* 78. Bandung: Nix.
- 1956 Verb structure in Javanese. In: M. Halle et al., comps *For Roman Jakobson*, 567-573. The Hague: Mouton.
- 1975 Sentence segment and word group: basic concepts of Javanese syntax. In: J.W.M. Verhaar, ed. *Linguistic studies in Indonesian and languages in Indonesia* 1:6-10. Jakarta: NUSA.

SYNCHRONICAL DESCRIPTION AT THE PHONETIC AND SYLLABIC LEVEL
OF MODANG (KALIMANTAN TIMUR) IN CONTRAST TO KENYAH, KAYAN,
AND PALAWAN (PHILIPPINES)

Nicole Revel-Macdonald

In 1974 a meeting on 'The Study of Malay Culture' was held in Bali and in accordance with a recommendation of the Advisory Committee of UNESCO, the Directorate General of Culture of the Republic of Indonesia in co-operation with the Indonesian National Commission for UNESCO convened a workshop in ethnomusicology, oral traditions and the performing arts¹ which took place in Kalimantan Timur from March 29 to April 23, 1977. The workshop was an itinerant one on the Mahakam Basin and two of its eastern tributaries, the Kelinyau and the Telen.

We lived on two taxi-boats for three weeks and the fieldwork covered a distance of 1,000 kms to and from Samarinda. Along the way, we visited six Dayak villages (cf. map 1) but for a very brief period of time ranging from two to three days, sometimes less.

Each village welcomed us in the most courteous and ceremonial manner and presented to us the main items of its repertoire of music and dances, along with which we were taping and learning from observations and dialogues — thanks to the co-operation of the villagers. In such intense days and nights of work very little time was left for systematic linguistic inquiries but, in the respective villages visited, I tried to fill out a list of vocabulary. I only failed once in Long Səgar — among the Kenyah Uma? Jalam — because of an anticipated departure.

In oriental Asia comparative linguistics has to rely on vocabularies rather than on syntax and morphology to identify the genuine features of the given languages in order to classify them.

A list of word bases considered unequivocal and belonging to the old stock of the Austronesian world was proposed at the beginning of the workshop in Jakarta; some addenda and substitutions were made in order to adapt it to the environment and technology of the people of Kalimantan. It was based on the CEDRASEMI ethno-linguistic atlas of ~ 800 words among which 109 were to be accorded priority.

The list of 109 wordbases is semantically organised (natural phenomena (7), metal (6), plants (18), animals (10), alimentation (4), parts of the body (13), tools (11), colours (3) and numbers (13)) and by observing the vicinity and

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 321-331. *Pacific Linguistics*, C-75, 1982.

© Nicole Revel-Macdonald

proportion of cognates and non-cognates we can see three main groups:

- The first groups we visited along the river were speakers of *Melayu Kutai* and *Melayu Banjar*, which share a very high range of cognates with *Bahasa Indonesia* (our starting point in the inquiry). We have one subgroup.

- The second subgroup consists of *Bugis* with 72 vocabulary items, among which 26 are non-cognates.

Məlayu Kutai, *Məlayu Banjar* and *Bugis* are the languages spoken by the Islamic people living as merchants and traders in the east coast of Kalimantan, in the main towns of Balikpapan, Samarinda and Tenggarong, and farther upstream in Muara Kaman, Muara Ancalong (cf. map 1) and the big villages at the junction of the rivers and their tributaries.

- The third group consists of three Dayak languages and their respective local dialects, namely:

Modang with the dialect of Long Bəntə[?] (Ma); the dialect of Nəhəs Liah Bing (Mb);
Kenyah with the dialect of Long Lə[?]əs (Ka), subgroup Uma[?] Bəm;
 the dialect of Tanjung Manis (Kb), subgroup Uma[?] Taw;
 the dialect of Long Noran (Kc), subgroup Ləpə[?] Kulit;
Kayan in Miau Baru on the Wahau River (Ky).

In this paper we shall observe how one of the Dayak languages — namely *Modang* — reacts to diverse factors of evolution, for it seems possible to discover a synchronic dynamism by comparing *Modang* to the neighbouring *Kenyah* and *Kayan* languages as well as to the two Malay dialects, *Banjar* and *Kutai*, and extending the comparison to a language of the southern Philippines (neighbouring Borneo) which we are familiar with, *Palawan*.

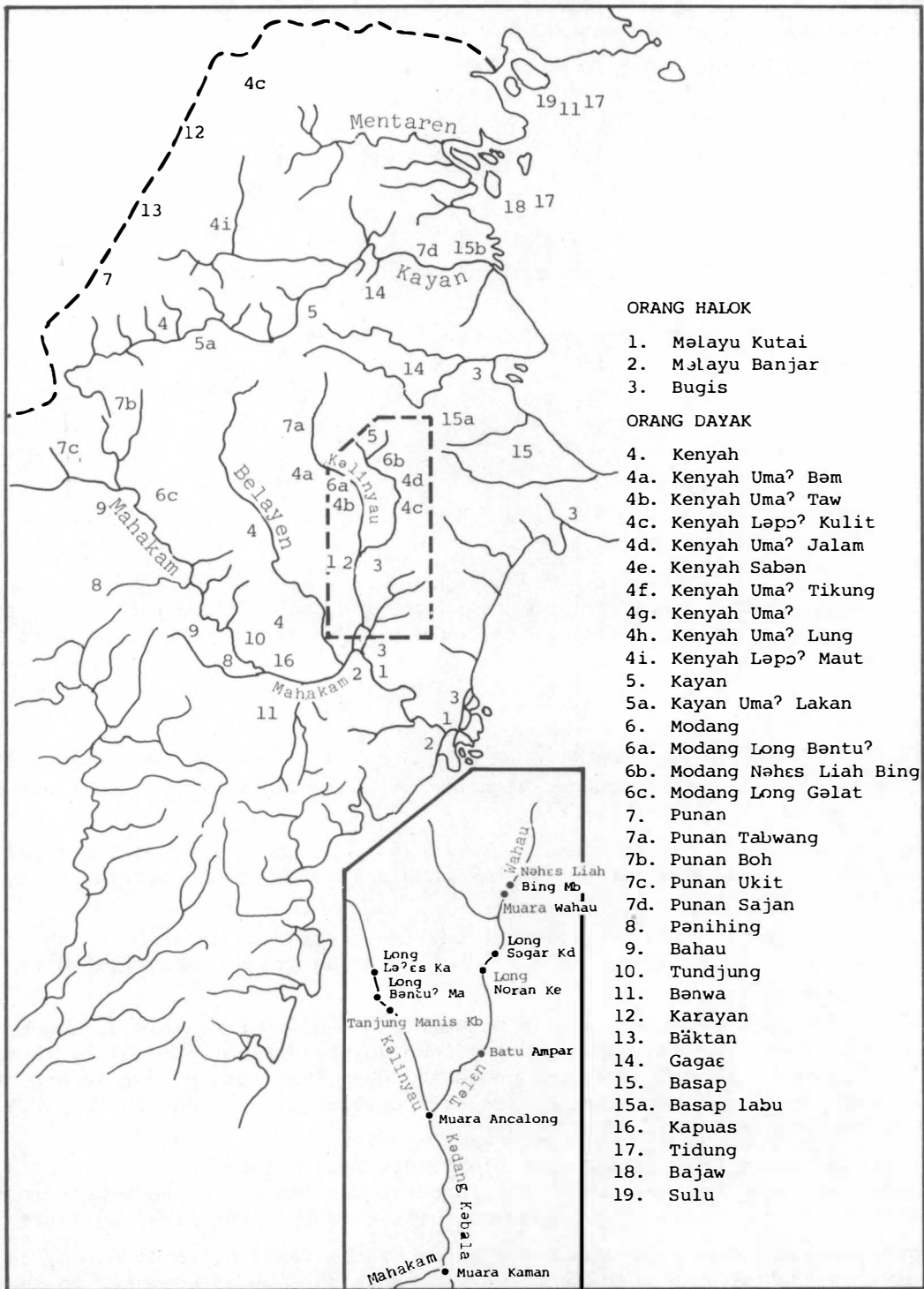
Instead of focussing our interest on the causes of the noticeable evolution of *Modang* — causes which are not only physical or physiological but also and necessarily extralinguistical because of migrations, bilingualism, borrowings, contemporary modernisation, propagation of various faiths, and education — we shall proceed to the phonetic description of *Modang* and its syllabic structure in an attempt to describe, by reference to our modest documentation, the tendencies of this language. As a matter of fact it is characterised by a marked tendency to move from di- or tri-syllabism to monosyllabism.

SYNCHRONIC STUDY OF THE SYLLABIC STRUCTURE OF MODANG (Mb)

A comparison between the *Modang*, *Kenyah* and *Kayan* languages as well as the different Malay dialects and *Bugis* shows a clear tendency to reduce the disyllables to monosyllables and trisyllables to disyllables or even monosyllables.

This reduction is operated in the initial syllable(s):

- | | | | | |
|----|--------------|----------------|-------------------|------------------|
| 1. | (V)CVC | : umah (Kb) | → mə [?] | 'dry rice field' |
| | | ənam (BAH) | → nam | 'six' |
| | | əmas (BAH) | → maes | 'gold' |
| 2. | (VC)CVC | : udjan (BAH) | → si:n | 'rain' |
| 3. | C(V)CVC | : malat (Kb) | → mlət | 'iron' |
| 4. | (CV)CVC | : ləsong (BAH) | → sɔŋ | 'mortar' |
| 5. | (CV) (CV)CVC | : kələput (Kb) | → pət | 'blowpipe' |
| 6. | CV(C) | : asaw (Ka) | → sa:° | 'dog' |



Map 1: Listing and location of languages in East Kalimantan

Consequently the weight passes from the beginning to the ending of the word and manifestation of it has several aspects:

a) Lengthening of the vowel in an open syllable of a monosyllabic or disyllabic word	CV ²	lɔ: lədzə:	'betel leaf' 'wood'
b) Consonant followed by a diphthong in an open syllable of a monosyllabic or disyllabic word	Cdiphthong	psæɛ ləkɔɛ	'paddle' 'man'
c) Absence of lengthening of the vowels in a closed monosyllable or disyllable	CVC	sɔŋ	'mortar'
d) Consonant followed by a diphthong in a closed syllable	CdiphthongC	tɔŋbæɛs məlhɛʔŋ	'jew's harp' 'red'
e) Succession of two vowels in open monosyllabic word	CCV	pláɛ ptáɛ séɛ	'rice' 'banana' 'adze'
f) Succession of two vowels in a closed monosyllabic word following a metathesis	CCVVC	tluang	'bone'

In other words the contraction or, if you wish, the erasing of the first syllable in Modang (Ma) and (Mb) is followed by a continuity in the final syllable which might be:

- lengthening of the vowel
- diphthongs
- succession of two vowels.

Obviously this is a language which phonetically manifests a high range of possibilities in the final syllable.

The evolution of Modang seems to be linked with the weight of the final articulation. It is demonstrated at the phonic level by two characteristics, namely stress and length.

Stress being fixed is not relevant and the strong accentuation on the last syllable weakens the first syllable, and by so doing has probably tended to erase it completely.

There is no relevant opposition between long and short vowels although, at the phonetic level, we can perceive many long vowels articulated mainly in Ma. Stress and length are linked at the phonetic level, but length is relevant at the phonemic level and the canonic form of the vowel is not long in Mb, while it seems to be long in Ma (cf. chart 2).

The complexity in the last syllable of disyllables, or in monosyllables, manifested by lengthening, diphthong², and the succession of two vowels is to be carefully noted, being the consequence of the fall of the first syllable.

Therefore we proceeded to a brief study of the last rhymes of monosyllabic wordbases in the two Modang dialects we were able to tape as compared to Kenyah and Kayan, the neighbouring Dayak languages in this area.

This table of correspondences between cognates manifests the progressive variations on the three Dayak languages we are focussing upon, and their respective dialects.

The starting point is a simple known etymon consisting of a simple vowel and one can see the process of diphthongisation characteristic of Modang (Ma). The tables focus on Modang and should be read from Modang to Kenyah and Kayan.

The tendency towards a reduction of disyllables into monosyllables produced:

1) *Initial consonantic group consecutive to the loss of a vowel*

pl	:	paraj (Kc)	→	plæ	'rice'
bl	:	bulong (Ky)	→	blung	'sago'
tl	:	tulang (Kc)	→	tluang	'bone'
pt	:	pəttəj (Kc)	→	ptæ	'banana'
ps	:	pəsəj (Kc)	→	psa:ɛ	'fishhook'
ts	:	tusu (Kc)	→	tsu ^ʔ	'seven'
ml	:	malat (Kb)	→	m ^h læɪt	'iron'
mt	:	m ^h ytan (Mb)	→	m ^h tan	'eye'
ms	:	məsə (Ma)	→	m ^h sə ^ʔ	'house'
m ^h n	:	m ^h anək (Ky)	→	m ^h nək	'bird'

We observe that in each initial consonantic group the second consonant is always an apico or alveolar |C + t, s, l, n| the nasal |m + C| seem to be an initial consonantic group as the occlusives |p + C|, |t + C| or |b + C| and the group |m + apicoalveolar| enters in correlation with the former ones.

This rule of formation of initial consonantic groups is verified by the fact that with velar and palatals as a second consonant, the fall of the vowel which is very often a *pepet* is not operated:

Ma	kədjó	'tree'
	ləká:ɛ	'man'

2) *Implosive consonants*

In absolute final position after a vowel I perceive implosive consonants $\underset{\wedge}{t}$ $\underset{\wedge}{k}$ $\underset{\wedge}{p}$ as in Maanyan.

This tendency towards an implosive articulation after a final vowel is effective in the three orders.

3) Most probably the evolution of weakening of the initial bilabial consonant into a bilabial lightly velar continuant b > w like bulan > wəlwen 'moon', is also to be linked to the weight (stress and diphthong) of the syllable.

WORDBASE STRUCTURE

Everything works as if a transposition of disyllabism was operated on the remaining syllable. Consequently and necessarily the basic form of the syllabic nucleus is complex (long vowels in Ma; diphthongs in Mb or a succession of two vowels).

The wordbase structure is either disyllabic or monosyllabic. In the last case we observed several phenomena at the vowel level but also at the consonant level:

- 1) After a simple consonant the weight is at the vowel level manifested by the lengthening or the burst into diphthongs.
- 2) After the loss of the vowel of the first syllable an initial consonant cluster appears providing the second consonant is apico-alveolar.

This is apparently a constraint of the word structure.

	'gold'	'husked rice'	'four'	'fire'	'rice in the field'	'paddle'	'rat'	'eyes'	'iron'
Ma	mé:ah	hé:a	pé:a _ɿ	póε	pəllé:	pəsé:	əβéa	məté ^a n	məl ^a k
Mb	mæs	hæs	pət	əpɔ ^ε	pláε	psæ ^ε	əβéa	mtan	mlæt
Ka	mas	ba:h		apój	padáj	bəsó _j	bəlabaw	matá	utón
Kb	ma _ɿ	ba'á	pat	lutón	padój	bəsój	bəlabáw	matá	malát
Kc	mas	bahá	pat	apój	paráj	bəsáj	bəlabáw	maté	utón
	ma _ɿ	bahá	pat	apój	paráj	bəsáj	laβáw	matán	titáj
	PPH amas PAN hemas	PPH beRas	PPH epat	PAND 'apuj	PAND pag'aj	PPH beR(a)say	PPH la(bw)aw	PAND mata	

This chart tends to show the monosyllabic tendency of Modang and diphthong occurrence.

NB: Although PPH reconstructions are cited (see Wurm and Wilson 1975), that does not imply that the languages discussed here are in a Philippine subgroup. Most of these reconstructions are of a quarter time depth.

Chart 1

	'nose'	'mortar'	'dog'	'hair' (head)	'head'	'wood'	'hair' (body)	'to eat'	'seven'	'blow pipe'	'banana'	'stone'
Ma	lo:ŋ	so:ŋ	sa:°	wɔ:k̡	dɔ:g̡	kəɟo:	bəllo:n	mu:n	sóu	pó:t	pətó:a	wətá°
Mb	luŋ	sɔŋ	tloŋ	wɔk	doʔ	kəɟy:	blun	hoʔ	tsuʔ	pɔt	ptéɛ	wətáɔ
Ka	ndoŋ	ləsóŋ	asó	pó:k	uló	kajó	bulú	umán		kəlləpút	pəttih	batú
Kb	ŋdoŋ	lɔsoŋ	asú	pók	ulú	kajú	bulú	umán	tudzoʔ	kələpút	pəttí	batú
Kc	ŋtoŋ	lisóŋ	asow	buk	uláw	kajáw	buláw	umán	tusúʔ	kləpuṭ̡	pəttɔj	batáw
Ky	uróŋ	so:ŋ	asɔ:ʔ	bwok buk	kahon	kajóʔ	bulóʔ	kani	tusúʔ	hapɔṭ̡	putéʔ	batóʔ
	PAND ʔugʔung	PANS lěsun	PANS asu	PANDF buhək	PANC ulu	PAND kaju	PANDF bulu		PANDLRD tudʔuh	PAND tulup	PANS punti	PAND batuʔ

This chart tends to show the monosyllabic tendency of Modang on rhymes with [o] and [u].
At first approximation [o] and [u] are one single phoneme.

Chart 2

	'rain'	'tooth'	'hen'	'pig'	'five'	'swidden'	'man'	'fishhook'
Ma	si:n	ki:v	dzip _ˆ	dji:m	mɔ:ŋ	mɔ:g _ˆ	ləká:ɛ	pəsá:ɛ
Mb	sjən	kwo:	djəp	djəm	məʔ	məʔ	ləkɔ ^ɛ	gɪt
Ka	udján	dʷipén	ijap	baboj		umá	lakɪ	pa:sí
Kb	udján	dzipén	jap	buʔin	lɔmma	umáh	lake	kawit
Kc	usán	dzipén	ijap	baboj	ləmmé	uméh	lakəj	pəsəj
Ky	usá:n	ipén	hijap	baβuj	limaʔ	lumɔʔ	lɔkɛʔ	kawit _ˆ
	PAMS udan	PAND ipən	PANBLAA piak	PAND babuj	PAND lima	PTSL uma	PAND lakɪ	PANDLO kawit

This chart tends to show the monosyllabic tendency of Modang on rhymes with /i/, /ə/, and /ɔɔ/, /ə/

Chart 3

From charts 1, 2 and 3, the following diphthongs are present in Modang:

a^o : cf. 'dog', 'stone'

o:^a : cf. 'banana'

ɛ^a : cf. 'gold', 'eyes', 'husked rice'

o^u : cf. 'seven'

ɔ^ɛ : cf. 'man', 'five'

e^a : cf. 'rat', 'four'

THE PROBLEM OF ? AND h

Referring to the work of I. Dyen (1953, 1971) and of A.G. Haudricourt (1966) we see that the treatment of the uvular protophoneme *q which is still articulated in the Paiwan and Atayal languages of Taiwan is:

*q > ? in intervocalic and final position of Kenyah and Kayan,

ex. (53) : dara? 'blood'

ex. (28) : pu?un 'tree'

This treatment is similar in languages of the Philippines, while in the Indonesian languages of Java and Sumatra, in Malay, as well as in Cham and Rhade in Continental South-East Asia the treatment is:

q > h

ex. (53) : darah (Bjr); darah (Bah)

ex. (28) : puhun (Bjr); pohon (Mal Ku) (Bah).

The treatment of Məlayu Banjar and Məlayu Kutai is distinct from Kenyah and Kayan as I observed them in that area. With Modang h being a laryngeal, we have:

ex. (53) : laehae? < dara? where the first syllable vowel assimilates to that of the second.

PROBLEM OF THE OCCURRENCE OF GEMINATES IN KENYAH (Ka), (Kb), (Kc)

The occurrence is automatic after the schwa which can arise from an original short vowel (Zorc 1978, and personal communication)

ex. (14)	'banana'	pəttih pətti pəttəj	} *punti
----------	----------	---------------------------	----------

ex. (3)	'three'	təllaw (Kc)	*təlu
---------	---------	-------------	-------

ex. (5)	'five'	lɔmma (Ka) ləmme (Kb)	*lɔma
---------	--------	--------------------------	-------

ex. (23)	'rice-seedling'	bənni? (Ka)	*bənhiq
----------	-----------------	-------------	---------

ex. (43)	'woman'	ləttoh (Ka)	?
----------	---------	-------------	---

The geminates are very frequent in Bugis and Madurese. They appear also in Modang (Ma).

ex. (50)	'tongue'	təlla?	*dilaq
----------	----------	--------	--------

ex. (51)	'hair of body'	bəllə:n	*buluh
----------	----------------	---------	--------

ex. (52)	'bone'	təlluwən	*tuq(ə)lan
----------	--------	----------	------------

ex. (30)	'sago'	bəllə:ŋ	?
----------	--------	---------	---

ex. (43)	'woman'	ləddəh → əldəh	?
----------	---------	----------------	---

In general, internal phonetic evolution is slow in large insular areas like Formosa and Borneo.

If one compares the evolution between certain languages of Borneo and those of Java, one can see that the evolution in Kalimantan Timur is more internally motivated (i.e. unaffected by the extraneous influences of other languages) than in Java where Hindic influence affected the shape of words.

Some Dayak groups of Borneo have a typology very similar to the one observed in the Southern Philippines: the Murut (cf. Prentice 1969) and Dusun languages of Sabah are very close to the Palawan (cf. Revel-Macdonald 1978:282) language at the phonemic level and in syllabic structure. I would consider Kenyah and Kayan as having a parallel evolution with this branch of the Austronesian family (typologically not genetically).

In the areas of the Kelinyau, Telen and Wahau Rivers which we visited, we could observe that, in the middle of rather classic and normalised languages like Kenyah and Kayan, the Modang language was evolved more drastically. While the two are mutually intelligible, Modang is experienced as a strong linguistic boundary between groups.

It would be interesting to see what effectively happened to the vocalic system of Modang. The original system is a simple one consisting of four vowels:

i u
 ə
 a

The oppositions underlining the diphthongs are still to be brought out.

The 'burst' of the vowel in the last remaining syllable should be analysed in relationship to all the other phonemes and of course to the morphological patterns of this language.

One can raise the question as to how long it would take to develop the features described for Modang in these pages.

NOTES

1. See the UNESCO publication (1979).
2. In order to distinguish a diphthong from a succession of two vowels one hears two morae instead of three in the whole word.

BIBLIOGRAPHY

DYEN, Isidore

- 1953 *The Proto-Malayo-Polynesian laryngeals*. Baltimore: The Linguistic Society of America.
- 1971 The Austronesian languages and Proto-Austronesian. In: Thomas A. Sebeok, ed. *Current trends in linguistics*, vol.8:5-54.

HAUDRICOURT, A.-G.

- 1974 Limites et connections de l'Austro-asiatique au nord-est. *ASEMI* 5/1. Also in N. Ziede, 1966, *Studies in comparative Austroasiatic linguistics*.

PRENTICE, D.J.

- 1969 Phonemes of Sabah Murut. *Papers in Borneo linguistics* 1:23-41. *PL*, A-20.

REVEL-MACDONALD, Nicole

- 1978 *Le Palawan: phonologie, catégories, morphologie*. Paris: SELAF.

UNESCO

- 1979 *Music of the Kenyah and the Modang*. Quezon City: University of the Philippines. (J. Maceda, N. Revel-Macdonald, I Made Bandem, et al.)

WURM, S.A. and B. WILSON

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

ZORC, R. David

- 1978 Proto-Philippine word accent: innovation or Proto-Hesperonesian retention? In: S.A. Wurm and Lois Carrington, eds *Second International Conference on Austronesian Linguistics: proceedings*, 67-119. *PL*, C-61.

PACIFIC LINGUISTICS - LIST OF PUBLICATIONS

After each entry, ISBN numbers have been added. Where there is more than one number, the first refers to the complete set of volumes and the second and/or subsequent numbers to individual volumes or parts.

Note: the earliest works were called LINGUISTIC CIRCLE OF CANBERRA PUBLICATIONS, however all now bear the name PACIFIC LINGUISTICS.

All volumes are softbound unless otherwise indicated.

Unmarked publications which come under "The influence of English and other metropolitan languages in the Pacific area" project are: A-54, A-57, B-26, B-61, B-73, C-34, C-40, C-52, D-3, D-5, D-12, D-23, D-29.

SERIES A - OCCASIONAL PAPERS

- No. 1. WURM, S.A. Some remarks on the role of language in the assimilation of Australian aborigines. 1963; 12pp. Reprinted 1966. ISBN 0 85883 006 X
- No. 2. HEALEY, Alan Handling unsophisticated linguistic informants. 1964; iii+30pp. Reprinted 1967, 1972, 1973, 1975. ISBN 0 85883 007 8
- No. 3. Papers in New Guinea linguistics No.1. 1964; iv+42pp. Reprinted 1971. ISBN 0 85883 008 6
Papers by Alan Pence; Ellis Deibler Jr; Phyllis M. Healey; Bruce A. Hooley.
- No. 4. Papers in New Guinea linguistics No.2. 1964; iv+41pp. (incl. 1 map). Reprinted 1971. ISBN 0 85883 009 4
Two papers by S.A. Wurm.
- No. 5. Papers in New Guinea linguistics No.3. 1965; iv+53pp. Reprinted 1972. ISBN 0 85883 010 8
Two papers by Phyllis M. Healey.
- No. 6. Papers in New Guinea linguistics No.4. 1965; iv+68pp. Reprinted 1971. ISBN 0 85883 011 6
Two papers by Darlene Bee.
- No. 7. Papers in New Guinea linguistics No.5. 1966; viii+93pp. Reprinted 1971. ISBN 0 85883 012 4
Papers by Chester I. and Marjorie E. Frantz; Des and Jennifer Oatridge; Richard E. Loving; Joyce Swick; Alan Pence; Philip Staalsen; Helen and Maurice Boxwell.
- No. 8. Papers in Philippine linguistics No.1. 1966; iv+38pp. Reprinted 1971. ISBN 0 85883 013 2
Papers by Jo Shetler; Richard Pittman; Vivian Forsberg; Jean Hussey.
- No. 9. Papers in South East Asian linguistics No.1. 1967; iii+43pp. + 30 charts, 24 tables. Reprinted 1970. ISBN 0 85883 014 0
Papers by Nguyễn Đăng Liêm (2); A. Tran Huong Mai; David W. Dellinger.
- No.10. Papers in Australian linguistics No.1. 1967; v+59pp. Reprinted 1972. ISBN 0 85883 015 9
Papers by David and Kathleen Glasgow; Jean F. Kirton; W.J. Oates; B.A. and E.G. Sommer.
- No.11. Papers in Australian linguistics No.2. 1967; iii+73pp. + 7 maps. Reprinted 1971. ISBN 0 85883 016 7
Papers by C.G. von Brandenstein; A. Capell (2); Kenneth Hale.
- No.12. Papers in New Guinea linguistics No.6. 1967; iii+48pp. Reprinted 1971. ISBN 0 85883 017 5
Papers by K.A. McElhanon; G.L. Renck.
- No.13. Papers in New Guinea linguistics No.7. 1967; iv+59pp. Reprinted 1971. ISBN 0 85883 018 3
Papers by Jean Goddard; Karl J. Franklin.

Series A - Occasional Papers (continued)

- No.14. Papers in Australian linguistics No.3. 1968; iii+46pp. + 1 map.
Reprinted 1971. ISBN 0 85883 019 1
Papers by E.F. Aguas; D.T. Tryon.
- No.15. Papers in linguistics of Melanesia No.1. 1968; iii+52pp. + 1 map.
Reprinted 1971, 1980. ISBN 0 85883 020 5
Papers by A. Capell; G.J. Parker; A.J. Schütz (2).
- No.16. Papers in New Guinea linguistics No.8. 1968; iv+62pp. (incl. 2 maps).
Reprinted 1971. ISBN 0 85883 021 3
Papers by C.L. Voorhoeve; Karl J. Franklin; Graham Scott.
- No.17. Papers in Australian linguistics No.4. 1969; viii+97pp. (incl. 3 maps).
Reprinted 1971. ISBN 0 85883 022 1
Papers by Joy Kinslow Harris; S.A. Wurm; Don Laycock.
- No.18. Papers in New Guinea linguistics No.9. 1969; vi+110pp. (incl. 1 map).
Reprinted 1971. ISBN 0 85883 023 X
Papers by A. Capell; Alan Healey; Darryl Wilson (3).
- No.19. Papers in Philippine linguistics No.2. 1969; iii+32pp. Reprinted 1971.
ISBN 0 85883 024 8
Papers by Jeanne Miller; Helen W. Miller.
- No.20. Papers in Borneo linguistics No.1. 1969; iv+41pp. Reprinted 1971.
ISBN 0 85883 025 6
Papers by D.J. Prentice (3).
- No.21. Papers in linguistics of Melanesia No.2. 1969; v+105pp. (incl. 5 maps).
Reprinted 1979. ISBN 0 85883 002 7
Papers by A. Capell; Ann Chowning; S.A. Wurm.
- No.22. Papers in New Guinea linguistics No.10. 1969; v+84pp. ISBN 0 85883 026 4
Papers by Don Laycock; Richard G. Lloyd; Philip Staalsen.
- No.23. Papers in New Guinea linguistics No.11. 1970; v+78pp. (incl. 1 map).
ISBN 0 85883 001 9
Papers by Gordon and Ruth Bunn; Alan Pence, Elaine Geary and Doris Bjorkman; Harry and Natalia Weimer; O.R. Claassen and K.A. McElhanon.
- No.24. Papers in Philippine linguistics No.3. 1970; vi+77pp.
ISBN 0 85883 000 0
Papers by Norman Abrams; Jannette Forster; Robert Brichoux.
- No.25. Papers in New Guinea linguistics No.12. 1970; iv+60pp. + 1 map.
ISBN 0 85883 027 2
Papers by C.L. Voorhoeve; K.A. McElhanon; Bruce L. Blowers.
- No.26. Papers in New Guinea linguistics No.13. 1970; iv+48pp. ISBN 0 85883 028 0
Papers by Bruce L. Blowers; Margie Griffin; K.A. McElhanon.
- No.27. Papers in Australian linguistics No.5. 1971; iv+70pp. ISBN 0 85883 029 9
Two papers by Jean F. Kirton.
- No.28. Papers in New Guinea linguistics No.14. 1971; vi+172pp. (incl. 8 maps).
ISBN 0 85883 030 2
Papers by T.E. Dutton; C.L. Voorhoeve; S.A. Wurm.
- No.29. Papers in South East Asian linguistics No.2. 1971; iv+78pp. (incl. 1 map).
ISBN 0 85883 031 0
Papers by Warren W. Glover; Maria Hari; E.R. Hope.
- No.30. Papers in South East Asian linguistics No.3. 1973; iv+82pp.
ISBN 0 85883 091 4
Papers by D.W. Dellinger; E.R. Hope; Makio Katsura; Tatsuo Nishida.
- No.31. Papers in New Guinea linguistics No.15. 1972; v+69pp. ISBN 0 85883 032 9
Papers by R.K. Lewis; Sandra C. Lewis; Shirley Litteral; Philip Staalsen.
- No.32. Papers in Philippine linguistics No.4. 1971; iv+32pp. ISBN 0 85883 033 7
Papers by R.M. Hohulin; Lou Hohulin.

Series A - Occasional Papers (continued)

- No.33. Papers in Borneo and Western Austronesian linguistics No.2. 1977; vi+132pp. + 1 map. ISBN 0 85883 164 3
Papers by C. Court; Robert A. Blust; F.S. Watuseke.
- No.34. Papers in New Guinea linguistics No.16. 1972; iii+46pp. ISBN 0 85883 081 7
Papers by Janice Allen; Marshall Lawrence.
- No.35. Papers in linguistics of Melanesia No.3. 1972; vii+113pp. (incl. 6 maps) + 6 maps. ISBN 0 85883 083 3
Papers by C.H. Beaumont; D.T. Tryon; S.A. Wurm.
- No.36. Papers in Australian linguistics No.6. 1973; iv+72pp. + 4pp. photographs, 2 maps. ISBN 0 85883 095 7
Papers by B. Schebeck; Luise A. Hercus and Isobel M. White.
- No.37. Papers in Australian linguistics No.7. 1974; iv+73pp. (incl. 1 map, 3 photographs). ISBN 0 85883 116 3
Papers by Christine E. Furby; Luise A. Hercus; Christine Kilham.
- No.38. Papers in New Guinea linguistics No.17. 1973; iii+78pp. ISBN 0 85883 097 3
Papers by K.G. Holzkmnecht (3); Donald J. Phillips.
- No.39. Papers in Australian linguistics No.8. 1975; v+78pp. ISBN 0 85883 126 0
Papers by M.C. Sharpe; Lothar Jagst; David B.W. Birk.
- No.40. Papers in New Guinea linguistics No.18. 1975; iv+102pp. (incl. 5 maps). ISBN 0 85883 118 X
Papers by Robert Conrad and Wayne Dye; N.P. Thomson; Leslie P. Bruce Jr.
- No.41. Papers in Philippine linguistics No.5. 1974; iv+74pp. ISBN 0 85883 114 7
Papers by Donna Hettick Chandler; Edward Ruch; Jeannette Witucki.
- No.42. Papers in Australian linguistics No.9. 1976; iv+79pp. ISBN 0 85883 140 6
Papers by Joyce Hudson; Barbara J. Sayers.
- No.43. Papers in Philippine linguistics No.6. 1974; iii+74pp. (incl. 1 map). ISBN 0 85883 108 2
Papers by Thomas N. Headland and Alan Healey; Jeannette Witucki.
- No.44. Papers in Philippine linguistics No.7. 1975; iv+60pp. ISBN 0 85883 135 X
Papers by Betty Hooker; Dietlinde Behrens; Patricia M. Hartung.
- No.45. Papers in New Guinea linguistics No.19. 1976; v+105pp. (incl. 2 maps). ISBN 0 85883 156 2
Papers by Ger P. Reesink; Lillian Fleischmann and Sinikka Turpeinen; Peter C. Lincoln.
- No.46. Papers in Philippine linguistics No.8. 1976; iv+89pp. ISBN 0 85883 146 5
Papers by Jeannette Witucki; Michael R. Walrod; Jean Shand.
- No.47. Papers in Australian linguistics No.10. 1976; iv+78pp. (incl. 3 maps, 11 photographs). ISBN 0 85883 153 8
Papers by Jean F. Kirton; Bruce A. Sommer; S.A. Wurm and L. Hercus; P. Austin, R. Ellis and L. Hercus.
- No.48. THOMAS, David, Ernest W. LEE and NGUYỄN ĐĂNG LIÊM, eds
Papers in South East Asian linguistics No.4. Chamic studies. 1977; ix+124pp. ISBN 0 85883 163 5
Papers by Alice Tegenfeldt Mundhenk and Hella Goschnick; Timothy Friberg and Kvoeu Hor; Doris Walker Blood; David L. Blood; Eugene Fuller; Ernest W. Lee; Hella Goschnick.
- No.49. Papers in South-East Asian linguistics No.5. 1977; iv+98pp. ISBN 0 85883 158 9
Three papers by David Bradley.

Series A - Occasional Papers (continued)

- No.50 Papers in Philippine linguistics No.9. 1979; v+108pp. (incl. 1 map).
ISBN 0 85883 186 4
Papers by E. Clay Johnston; Hartmut Wiens; Jo Ann Gault with Sulaiman and Fatima Barhama; Peter Green; Bruce Grayden; Jeannette Witucki.
- No.51. Papers in Australian linguistics No.11. 1978; vii+199pp. (incl. 1 map) + 3 maps, 4 photographs. ISBN 0 85883 179 1
Papers by Jean F. Kirton; R. Wood; L.A. Hercus; Chester S. Street and Harry Palada Kulampurut; Dianne Buchanan; Jean F. Kirton and Bella Charlie.
- No.52. GETHING, T.W. and NGUYỄN ĐĂNG LIÊM, eds Papers in South-East Asian linguistics No.6: Tai studies in honour of William J. Gedney. 1979; vi+149pp. ISBN 0 85883 188 0
Papers by Pongsri Lekawatana; Leslie M. Beebe; Thomas W. Gething; Carol J. Compton; Mary Sarawit; Thomas Scovel; John F. Hartmann; Wilaiwan Khanittanan; James R. Chamberlain; Beatrice T. Oshika.
- No.53. Papers in South-East Asian linguistics No.7. 1980; v+130pp. (incl. 4 maps) ISBN 0 85883 206 2
Papers by Ronald L. Trail with Harisingh T. Rathod, Geeta Chand, Chaudhary Roy, Indira Shrestna, Nirmal Man Tuladhar; Peter J. Grainger; Warren W. Glover and John K. Landon; Austin Hale and Thakurlal Manandhar; Austin Hale; Burkhard Schöttelndreyer (3).
- No.54. Papers in pidgin and creole linguistics No.1. 1978; vi+197pp.
Reprinted 1980. ISBN 0 85883 178 3
Papers by Loreto Todd and Peter Mühlhäusler; S.A. Wurm (2); John T. Platt; Peter Mühlhäusler (2); D.S. Walsh.
- No.55. Papers in Philippine linguistics No.10. 1979; vi+142pp. ISBN 0 85883 193 7
Papers by Andrew F. Gallman; Joe E. Allison; Carol M. Harmon; Jeannette Witucki.
- No.56. Papers in New Guinea linguistics No.20. 1980; v+214pp. (incl. 4 maps). ISBN 0 85883 215 1
Papers by Maurice Boxwell; Jean Goddard; Malcolm Ross; Arden G. Sanders and Joy Sanders (2); Joy Sanders and Arden G. Sanders; H.J. Davies.
- No.57. Papers in pidgin and creole linguistics No.2. 1979; x+290pp. ISBN 0 85883 198 8
Papers by Peter Mühlhäusler; Elsa Lattey; Ellen Woolford; William G. Camden; Margaret S. Steffensen; M.G. Clyne; William Peet Jr; Ulrike Mosel; Ian Smith; Jerry G. Gebhard; Nguyễn Đăng Liêm; Gail Raimi Dreyfuss and Djoehana Oka; Maria Isabelita O. Dios de Riego.
- No.58. Papers in Australian linguistics No.12. 1980; vi+113pp. ISBN 0 85883 208 9
Papers by Chester S. Street (2); Helen Geytenbeek; Kathleen Glasgow and Mark Garner.
- No.59. RIGSBY, B. and P. SUTTON, eds Papers in Australian linguistics No.13: Contributions to Australian linguistics. 1980; viii+314pp. (incl. 5 maps). ISBN 0 85883 205 4
Papers by Alan Rumsey; Patrick McConvell; Peter Sutton (2); Tamsin Donaldson; L. Hercus; Peter Austin, Corinne Williams and Stephen Wurm; Paul Black; Terry Crowley; Harold J. Koch; D.T. Tryon; A.K. Chase and J.R. von Sturmer.
- No.60. Papers in Australian linguistics No.14. 1980; v+178pp. (incl. 33 maps). ISBN 0 85883 230 5
Papers by Bruce E. Waters (2); Peter A. Busby.
- No.61. Papers in New Guinea linguistics No.21. 1981; v+209pp. (incl. 10 maps). ISBN 0 85883 236 4
Papers by H.J. Davies; Dieter Osmers; John Lynch; S.A. Wurm.

Series A - Occasional Papers (continued)

- No.62. BRADLEY, David, ed. Papers in South-East Asian linguistics No.8: tonation. 1982; viii+158pp. (incl 1 map). ISBN 0 85883 256 9
Papers by A.V. Diller; David Bradley; Philip John Rose; Vũ Thanh Phương;
U Thein Tun; Jack and Mary Jane Gandour.

IN PREPARATION

Papers in New Guinea linguistics No.22.

Papers by Thomas R. Phinnemore; J. Miedema and F.I. Wellington; Malcolm Ross;
W.A.L. Stokhof and Don A.L. Flassy; J. Davies and B. Comrie; Hiroko Oguri;
Karen Adams and Linda Lauck; David Scorza; S. Abbott.

Papers in linguistics of Melanesia No.4.

Papers by John Lynch and others.

Papers in pidgin and creole linguistics No.3.

Papers by Allan Rumsey; Linda Simons; Jeff Siegel, Lois Carrington
and Peter Mühlhäusler.

Papers in Australian linguistics No.15.

Papers in Philippine linguistics No.11.

SERIES B - MONOGRAPHS

- No. 1. WURM, S.A. and J.B. HARRIS Police Motu: an introduction to the trade language of Papua (New Guinea) for anthropologists and other fieldworkers. 1963; vi+81pp. Reprinted 1964, 1965, 1966, 1967, 1969, 1970, 1971, 1973. ISBN 0 85883 034 5
- No. 2. WURM S.A. Phonological diversification in Australian New Guinea highlands languages. 1964; iii+87pp. + 1 map. Reprinted 1971. ISBN 0 85883 035 3
- No. 3. HEALEY, Alan Telefol phonology. 1964; ii+53pp. + 2 figures, 5 tables. Reprinted 1972, 1981. ISBN 0 85883 036 1
- No. 4. HEALEY, Phyllis M. Telefol noun phrases. 1965; iii+51pp. Reprinted 1972. ISBN 0 85883 037 X
- No. 5. HEALEY, Phyllis M. Levels and chaining in Telefol sentences. 1966; iv+64pp. Reprinted 1971. ISBN 0 85883 038 8
- No. 6. TRYON, Darrell T. Nengone grammar. 1967; x+91pp. Reprinted 1971. ISBN 0 85883 039 6
- No. 7. TRYON, D.T. Dehu grammar. 1968; xi+111pp. Reprinted 1971. ISBN 0 85883 040 X
- No. 8. TRYON, Darrell T. Iai grammar. 1968; xii+125pp. Reprinted 1971. ISBN 0 85883 041 8
- No. 9. DUTTON, T.E. The peopling of Central Papua: some preliminary observations. 1969; viii+182pp. Reprinted 1970, 1971. ISBN 0 85883 042 6
- No.10. FRANKLIN, K.J. The dialects of Kewa. 1968; iv+72pp. (incl. 20 maps). Reprinted 1971. ISBN 0 85883 043 4
- No.11. SOMMER, B.A. Kunjen phonology: synchronic and diachronic. 1969; iv+72pp. (incl. 3 maps). ISBN 0 85883 044 2
- No.12. KLOKEID, Terry J. Thargari phonology and morphology. 1969; viii+56pp. Reprinted 1981. (incl. 1 map). ISBN 0 85883 045 0
- No.13. TREFRY, D. A comparative study of Kuman and Pawaian. 1969; v+94pp. (incl. 1 map). Reprinted 1980. ISBN 0 85883 046 9
- No.14. McELHANON, K.A. Selepet phonology. 1970; v+47pp. (incl. 1 map). ISBN 0 85883 003 5
- No.15. TRYON, D.T. An introduction to Maranungku (Northern Australia). 1970; x+111pp. (incl. 1 map). ISBN 0 85883 047 7
- No.16. McELHANON, K.A. and C.L. VOORHOEVE The Trans-New Guinea Phylum: explorations in deep-level genetic relationships. 1970; v+107pp. (incl. 4 maps). Reprinted 1978. ISBN 0 85883 048 5
- No.17. KUKI, Hiroshi Tuamotuan phonology. 1970; ix+119pp. + 2 maps. ISBN 0 85883 049 3
- No.18. YOUNG, R.A. The verb in Bena-bena: its form and function. 1971; v+68pp. ISBN 0 85883 050 7
- No.19. PATON, W.F. Ambrym (Lonwolwol) grammar. 1971; xi+128pp. (incl. 1 map). ISBN 0 85883 051 5
- No.20. CAPELL, A. Arosi grammar. 1971; iv+90pp. (incl. 1 map). ISBN 0 85883 052 3
- No.21. McELHANON, K.A. Selepet grammar. Part I: From root to phrase. 1972; vi+116pp. ISBN 0 85883 085 X and 0 85883 086 8
- No.22. McELHANON, K.A. Towards a typology of the Finisterre-Huon languages, New Guinea. 1973. vii+73pp. (incl. 1 map). ISBN 0 85883 094 9
- No.23. SCOTT, Graham Higher levels of Fore grammar. Edited by Robert E. Longacre. 1973; x+88pp. ISBN 0 85883 088 4

Series B - Monographs (continued)

- No.24. DUTTON, T.E. A checklist of languages and present-day villages of central and south-east mainland Papua. 1973; iv+80pp. (incl. 1 map). ISBN 0 85883 090 6
- No.25. LAYCOCK, D.C. Sepik languages - checklist and preliminary classification. 1973; iv+130pp. (incl. 1 map). ISBN 0 85883 084 1
- No.26. MÜHLHÄUSLER, P. Pidginization and simplification of language. 1974; v+161pp. Reprinted 1978, 1981. ISBN 0 85883 113 9
- No.27. RAMOS, Teresita V. The case system of Tagalog verbs. 1974; viii+168pp. ISBN 0 85883 115 5
- No.28. WEST, Dorothy Wojokeso sentence, paragraph, and discourse analysis. Edited by Robert E. Longacre. 1973; x+181pp. ISBN 0 85883 089 2
- No.29. ELBERT, Samuel H. Puluwat grammar. 1974; v+137pp. ISBN 0 85883 103 1
- No.30. METCALFE, C.D. Bardi verb morphology (northwestern Australia). 1975; x+215pp. (incl. 1 map). ISBN 0 85883 121 X
- No.31. VOORHOEVE, C.L. Languages of Irian Jaya: checklist. Preliminary classification, language maps, wordlists. 1975; iv+129pp. (incl. 17 maps). Reprinted 1980. ISBN 0 85883 128 7
- No.32. WALTON, Janice Binongan Itneg sentences. 1975; vi+70pp. ISBN 0 85883 117 1
- No.33. GUY, J.B.M. A grammar of the northern dialect of Sakao. 1974; ix+99pp. (incl. 2 maps). ISBN 0 85883 104 X
- No.34. HOPE, Edward Reginald The deep syntax of Lisu sentences: a transformational case grammar. 1974; viii+184pp. (incl. 1 map). ISBN 0 85883 110 4
- No.35. IRWIN, Barry Salt-Yui grammar. 1974; iv+151pp. ISBN 0 85883 111 2
- No.36. PHILLIPS, Donald J. Wahgi phonology and morphology. 1976; x+165pp. ISBN 0 85883 141 4
- No.37. NGUYỄN ĐĂNG LIÊM Cases, clauses and sentences in Vietnamese. 1975; v+89pp. ISBN 0 85883 133 3
- No.38. SNEDDON, J.N. Tondano phonology and grammar. 1975; viii+264pp. ISBN 0 85883 125 2
- No.39. LANG, Adrienne The semantics of classificatory verbs in Enga (and other Papua New Guinea languages). 1975; xii+234pp. (incl. 2 maps). ISBN 0 85883 123 6
- No.40. RENCK, G.L. A grammar of Yagaria. 1975; xiii+235pp. (incl. 1 map). ISBN 0 85883 130 9
- No.41. Z'GRAGGEN, John A. The languages of the Madang District, Papua New Guinea. 1975; vi+154pp. (incl. 1 map). Reprinted 1979. ISBN 0 85883 134 1
- No.42. FURBY, E.S. and C.E. FURBY Preliminary analysis of Garawa phrases and clauses. 1977; viii+101pp. ISBN 0 85883 151 1
- No.43. STOKHOF, W.A.L. Preliminary notes on the Alor and Pantar languages (East Indonesia). 1975; vi+73pp. (incl. 2 maps). Reprinted 1979. ISBN 0 85883 124 4
- No.44. SAYERS, Barbara J. The sentence in Wik-Muncan: a description of propositional relationships. 1976; xvii+185pp. ISBN 0 85883 138 4
- No.45. BIRK, D.B.W. The MalakMalak language, Daly River (Western Arnhem Land). 1976; xii+179pp. (incl. 1 map). ISBN 0 85883 150 3
- No.46. GLISSMEYER, Gloria A tagmemic analysis of Hawaii English clauses. 1976; viii+149pp. ISBN 0 85883 142 2

Series B - Monographs (continued)

- No.47. SCOTT, Graham The Fore language of Papua New Guinea. 1978; xv+210pp. (incl. 2 maps). ISBN 0 85883 173 2
- No.48. CLARK, Marybeth Coverbs and case in Vietnamese. 1978; xi+215pp. ISBN 0 85883 162 7
- No.49. FILBECK, David T'in: a historical study. 1978; vi+111pp. (incl. 2maps). ISBN 0 85883 172 4
- No.50. SMITH, Kenneth D. Sedang grammar; phonological and syntactic structure. 1979; xix+191pp. (incl. 3 maps). ISBN 0 85883 180 5
- No.51. WELLS, Margaret A. Siroi grammar. 1979; vii+218pp. ISBN 0 85883 181 3
- No.52. KILHAM, Christine A. Thematic organization of Wik-Munkan discourse. 1977; xix+280pp. (incl. 1 map). ISBN 0 85883 168 6
- No.53. VESALAINEN, Olavi and Marja VESALAINEN Clause patterns in Lhomi. 1980; vii+100pp. ISBN 0 85883 210 0
- No.54. SNEDDON, J.N. Proto-Minahasan: phonology, morphology and wordlist. 1978; x+204pp. (incl. 1 map). ISBN 0 85883 169 4
- No.55. LYNCH, John A grammar of Lenakel. 1978; vii+135pp. (incl. 1 map). ISBN 0 85883 166 X
- No.56. ROSS, Malcolm with John Natu PAOL A Waskia grammar sketch and vocabulary. 1978; v+119pp. ISBN 0 85883 174 0
- No.57. BLAKE, Barry J. A Kalkatungu grammar. 1979; xii+198pp. (incl. 1 map). ISBN 0 85883 197 X
- No.58. BEAUMONT, Clive H. The Tigak language of New Ireland. 1979; xi+163pp. (incl. 2 maps). ISBN 0 85883 187 2
- No.59. STOKHOF, W.A.L. Woisika II: phonemics. 1979; xi+188pp. (incl. diagrams, photographs, 3 maps). Reprinted 1981. ISBN 0 85883 190 2
- No.60. FOX, G.J. Big Nambas grammar. 1979; xii+139pp. (incl. 2 maps). ISBN 0 85883 183 X
- No.61. HAWKINS, Emily A. Hawaiian sentence structures. 1979; iii+111pp. ISBN 0 85883 195 3
- No.62. HEATH, Jeffrey Basic materials in Ritharngu: grammar, texts and dictionary. 1980; ix+249pp. (incl. 1 map). ISBN 0 85883 204 6
- No.63. LUZARES, Casilda Edrial The morphology of selected Cebuano verbs: a case analysis. 1979; xii+208pp. ISBN 0 85883 199 6
- No.64. VOORHOEVE, C.L. The Asmat languages of Irian Jaya. 1980; x+177pp. (incl. 5 maps). ISBN 0 85883 207 0
- No.65. McDONALD, M. and S.A. WURM Basic materials in Waqumara (Galali): grammar, sentences and vocabulary. 1979; ix+111pp. (incl. 2 maps). ISBN 0 85883 202 X
- No.66. WOOLFORD, Ellen B. Aspects of Tok Pisin grammar. 1979; v+118pp. ISBN 0 85883 203 8
- No.67. HERCUS, L.A. The Bāgandji language. 1982; xviii+329pp. (incl. 5 maps, 10 photographs). ISBN 0 85883 263 1
- No.68. DAVIES, H.J. Kobon phonology. 1980; v+80pp. (incl. 3 maps). ISBN 0 85883 211 9
- No.69. SOBERANO, Rosa The dialects of Marinduque Tagalog. 1980; xii+232pp. (incl. 42 maps). ISBN 0 85883 216 X
- No.70. JOHNSTON, Raymond Leslie Nakanai of New Britain: the grammar of an Oceanic language. 1980; xiii+310pp. (incl. 2 maps). ISBN 0 85883 209 7
- No.71. CHAN YAP, Gloria Hokkien Chinese borrowings in Tagalog. 1980; viii+155pp. ISBN 0 85883 225 9
- No.72. HEATH, Jeffrey Basic materials in Warndarang: grammar, texts and dictionary. 1980; xii+174pp. (incl. 1 map). ISBN 0 85883 219 4

Series B - Monographs (continued)

- No.73. MOSEL, Ulrike Tolai and Tok Pisin: the influence of the substratum on the development of New Guinea Pidgin. 1980; viii+146pp. ISBN 0 85883 229 1
- No.74. WILLIAMS, Corinne J. A grammar of Yuwaalaraay. 1980; viii+200pp. (incl. 1 map). ISBN 0 85883 221 6
- No.75. GUY, J.B.M. Experimental glottochronology: basic methods and results. vii+217pp. ISBN 0 85883 220 8
- No.76. PERCIVAL, W.K. A grammar of the urbanised Toba-Batak of Medan. 1981; vi+125pp. ISBN 0 85883 237 2
- No.78. TSUNODA, Tasaku The Djaru language of Kimberley, Western Australia. 1981; xxi+290pp. (incl. 3 maps, 5 photographs). ISBN 0 85883 252 6
- No.79. GUY, J.B.M. Glottochronology without cognate recognition. 1981; viii+134pp. ISBN 0 85883 235 6
- No.80. DURANTI, Alessandro The Samoan fono: a sociolinguistic study. 1981; xi+195pp. (incl. 3 maps, 5 photographs). ISBN 0 85883 248 8
- No.81. TCHEKHOFF, Claude Simple sentences in Tongan. 1981; iv+95pp. ISBN 0 85883 251 8
- No.82. TIPTON, Ruth A. Nembu procedural and narrative discourse. 1982; v+87pp. ISBN 0 85883 259 3
- No.83. THURSTON, William R. A comparative study of Anêm and Lusi. 1982; ix+107pp. (incl. 3 maps). ISBN 0 85883 260 7
- No.85. WILSON, William H. Proto-Polynesian possessive marking. 1982; xv+137pp. ISBN 0 85883 270 4
- No.86. RUMSEY, A. An intra-sentence of Ngarinjin, North Western Australia. 1982; xii+179pp. (incl. 1 map) ISBN 0 85883 272 0

IN PREPARATION:

- HERCUS, Luise A. The languages of Victoria: a late survey.
- AUSTIN, Peter, Luise A. HERCUS, and Stephen A. WURM Basic materials in Malyangaba: grammar, sentences and vocabulary.
- CAUGHLEY, Ross The verb in Chepang.
- CROWLEY, Terry The Paamese language of Vanuatu.
- HARTMANN, John F. Linguistic and memory structures in Tai-lue oral narratives.
- McELHANON, K.A. The languages of the Morobe Province: checklist, classification and field guide.
- VERHEIJEN, J.A.J. The Sama/Bajan language of the Lesser Sunda Islands.
- MERLAN, Francesca Ngalakan grammar, texts and vocabulary.

NOTE: For further monographs on Indonesian languages see under Series D - the sub-series Materials in languages of Indonesia.

SERIES C - BOOKS

- No. 1. LAYCOCK, D.C. The Ndu language family (Sepik District, New Guinea). xi+224pp. (incl. 1 map). ISBN 0 85883 053 1
- No. 2. GRACE, George W. Canala dictionary (New Caledonia). 1975; ix+128pp. ISBN 0 85883 122 8
- No. 3. NGUYỄN ĐĂNG LIÊM English grammar: a combined tagmemic and transformational approach. (A contrastive analysis of English and Vietnamese, vol.1.). 1966; xliv+177pp. Reprinted 1970. ISBN 0 85883 054 X and ISBN 0 85883 055 8
- No. 4. NGUYỄN ĐĂNG LIÊM Vietnamese grammar: a combined tagmemic and transformational approach. (A contrastive analysis of English and Vietnamese, vol.2.). 1969; xlvi+209pp. Reprinted 1975. ISBN 0 85883 054 and ISBN 0 85883 056 6
- No. 5. NGUYỄN ĐĂNG LIÊM A contrastive grammatical analysis of English and Vietnamese. (A contrastive analysis of English and Vietnamese, vol.3.). 1967; xv+151pp. Reprinted 1971. ISBN 0 85883 054 X and ISBN 0 85883 057 4
- No. 6. TRYON, Darrell T. Dehu-English dictionary. 1967; v+137pp. Reprinted 1971. ISBN 0 85883 058 2
- No. 7. TRYON, Darrell T. English-Dehu dictionary. 1967; iii+162pp. Reprinted 1971. ISBN 0 85883 059 0
- No. 8. NGUYỄN ĐĂNG LIÊM A contrastive phonological analysis of English and Vietnamese. (A contrastive analysis of English and Vietnamese, vol.4.). 1970; xv+206pp. ISBN 0 85883 054 X and ISBN 0 85883 004 3
- No. 9. TRYON, D.T. and M.-J. DUBOIS Nengone dictionary. Part I: Nengone-English. 1969; vii+445pp. (out of print) ISBN 0 85883 060 4 and ISBN 0 85883 061 2
- No.10. OATES, W. and L. OATES Kapau pedagogical grammar. 1968; v+178pp. Reprinted 1971. ISBN 0 85883 062 0
- No.11. FOX, C.E. Arosi-English dictionary. 1970; iv+406pp. (incl. 1 map). ISBN 0 85883 063 9
- No.12. GRACE, George W. Grand Couli dictionary (New Caledonia). 1976; vii+113pp. ISBN 0 85883 154 6
- No.13. WURM, S.A. and D.C. LAYCOCK, eds Pacific linguistic studies in honour of Arthur Capell. 1970; xi+1292pp. (Hardbound: incl. 25 maps, 1 photograph). Reprinted 1974, 1975, 1978, 1979. ISBN 0 85883 005 1
Articles by Byron W. Bender, Catherine H. Berndt, Ronald M. Berndt, H. Bluhme, J.E. Bolt, C.G. von Brandenstein, C. Douglas Chrétien, J.R. Cleverly, Christopher Court, R.M.W. Dixon, Wilfrid H. Douglas, T.E. Dutton, Isidore Dyen, Samuel H. Elbert, A.P. Elkin, E.H. Flint, Karl J. Franklin, Marie Godfrey, George W. Grace, Kenneth Hale, Joy Harris, Alan Healey, Henry Hershberger, Ruth Hershberger, Patrick W. Hohepa, Nils M. Holmer, B.A. Hooley, Dorothy J. James, Hans Kähler, Susan Kaldor, Harland Kerr, Jean F. Kirton, D.C. Laycock, K.A. McElhanon, Howard McKaughan, Nguyễn Đăng Liêm, Geoffrey N. O'Grady, Andrew Pawley, Eunice V. Pike, Richard Pittman, D.J. Prentice, Albert J. Schutz, M.C. Sharpe, W.E. Smythe, A.J. Taylor, D.T. Tryon, E.M. Uhlenbeck, C.F. Voegelin, F.M. Voegelin, C.L. Voorhoeve, S.A. Wurm, John A. Z'graggen.
- No.14. GEERTS, P. 'Āre'āre dictionary. 1970; iv+187pp. (incl. 2 maps) ISBN 0 85883 064 7
- No.15. McELHANON, K.A. and N.A. McELHANON Selepet-English dictionary. 1970; xxi+144pp. ISBN 0 85883 065 5
- No.16. FRANKLIN, K.J. A grammar of Kewa, New Guinea. 1971; ix+138pp. ISBN 0 85883 066 3
- No.17. PARKER, G.J. Southeast Ambrym dictionary. 1971; xiii+60pp. ISBN 0 85883 067 1

Series C - Books (Continued)

- No.18. PRENTICE, D.J. The Murut languages of Sabah. 1971; xi+311pp.
(incl. 1 map). ISBN 0 85883 068 X
- No.19. Z'GRAGGEN, J.A. Classificatory and typological studies in languages of the Madang District. 1971; viii+179pp. (incl. 4 maps). ISBN 0 85883 069 8
- No.20. LANG, Adrienne Enga dictionary, with English index. 1973; lxi+219pp.
(Hardbound: incl. 1 map). Reprinted 1978. ISBN 0 85883 093 0
- No.21. PATON, W.F. Ambrym (Lonwolwol) dictionary. 1973; ix+337pp.
(Hardbound) + 1 map. ISBN 0 85883 092 2
- No.22. LONGACRE, Robert E., ed. Philippine discourse and paragraph studies in memory of Betty McLachlin. 1971; xv+366pp. (incl. 1 photograph).
ISBN 0 85883 070 1
Articles by Barbara Blackburn, R.E. Longacre, Betty McLachlin, Charles Walton, Claudia Whittle, Hazel J. Wrigglesworth.
- No.23. TRYON, D.T. and M.-J. DUBOIS Nengone dictionary. Part II: English-Nengone. 1971; iii+202pp. ISBN 0 85883 060 4 and ISBN 0 85883 071 X
- No.24. ELBERT, Samuel H. Puluwat dictionary. 1972; ix+401pp. (Hardbound).
ISBN 0 85883 082 5
- No.25. FOX, Charles E. Lau dictionary, with English index. 1974; vi+260pp.
(Hardbound). Reprinted 1976, 1978. ISBN 0 85883 101 5
- No.26. FRANKLIN, Karl, ed. The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea. 1973; x+597pp. (Hardbound: incl. 8 maps).
Reprinted 1975. ISBN 0 85883 100 7
Articles by H.A. Brown, T.E. Dutton, Karl J. Franklin, Richard G. Lloyd, George E. MacDonald, Karen Shaw, R. Daniel Shaw, Clemens L. Voorhoeve, S.A. Wurm.
- No.27. SOHN, Ho-min and B.W. BENDER A Ulithian grammar. 1973; xv+398pp.
(Hardbound: incl. 2 maps). Reprinted 1980. ISBN 0 85883 098 1
- No.28. HEADLAND, Thomas N. and Janet D. HEADLAND A Dumagat (Casiguran)-English dictionary. 1974; lxiii+232pp. (Hardbound: incl. 1 map).
ISBN 0 85883 107 4
- No.29. DUTTON, T.E., ed. Studies in languages of Central and South-East Papua. 1975; xvii+834pp. (Hardbound: incl. 5 maps). Reprinted 1978.
ISBN 0 85883 119 8
Articles by John Austing, Russell E. Cooper, T.E. Dutton, Cynthia Farr, James Farr, Roger Garland, Susan Garland, J.E. Henderson, J.A. Kolia, Mike Olson, Andrew Pawley, Ernest L. Richert, N.P. Thomson, Randolph Upia, Harry Weimer, Natalia Weimer.
- No.30. LOVING, Richard and Aretta LOVING Awa dictionary. 1975; xlv+203pp.
(Hardbound: incl. 1 map). ISBN 0 85883 137 6
- No.31. NGUYỄN ĐĂNG LIÊM, ed. South-East Asian linguistic studies, vol.1. 1974; vii+213pp. (Hardbound). Reprinted 1978. ISBN 0 85883 144 9 and
ISBN 0 85883 099 X
Articles by Marybeth Clark, Arthur G. Crisfield, Soenjono Dardjowidjojo, Cesar A. Hidalgo, Philip N. Jenner, Nguyễn Đăng Liêm, Saveros Pou.
- No.32. TRYON, D.T. Daly Family languages, Australia. 1974; xvii+305pp.
(Hardbound: incl. 1 map). Reprinted 1980. ISBN 0 85883 106 6
- No.33. WURM, S.A. and B. WILSON English finderlist of reconstructions in Austronesian languages (post-Brandstetter). 1975; xxxii+246pp.
(Hardbound). Reprinted 1978. ISBN 0 85883 129 5
- No.34. GUY, J.B.M. Handbook of Bichelamar - Manuel de Bichelamar. 1974;
iii+256pp. (Hardbound). Reprinted 1975, 1979. ISBN 0 85883 109 0
- No.35. KEESING, R.M. Kwaio dictionary. 1975; xxxv+296pp. (Hardbound: incl. 1 map). Reprinted 1981. ISBN 0 85883 120 1

Series C - Books (continued)

- No.36. REID, Lawrence A. Bontok-English dictionary. 1976; xxiii+500pp. (Hardbound). ISBN 0 85883 145 7
- No.37. RENCK, G.L. Yagaria dictionary, with English index. 1977; xxix+327pp. (Hardbound: incl. 1 map). ISBN 0 85883 161 9
- No.38. WURM, S.A., ed. New Guinea area languages and language study, vol.1: Papuan languages and the New Guinea linguistic scene. 1975; xlvi+1038pp. (Hardbound: incl. 28 maps). Reprinted 1977. ISBN 0 85883 131 7 and ISBN 0 85883 132 5
Articles by A. Capell, T.E. Dutton, Karl J. Franklin, Harland B. Kerr, D.C. Laycock, K.A. McElhanon, Evelyn M. Todd, C.L. Voorhoeve, S.A. Wurm, John A. Z'graggen.
- No.39. WURM, S.A., ed. New Guinea area languages and language study, vol.2: Austronesian languages. 1976; xxxv+736pp. (Hardbound: incl. 21 maps). Reprinted 1979. ISBN 0 85883 131 7 and ISBN 0 85883 155 4
Articles by C.H. Beaumont, A. Capell, Ann Chowning, T.E. Dutton, George W. Grace, Alan Healey, Bruce A. Hooley, D.C. Laycock, Peter C. Lincoln, David R. Lithgow, Andrew Pawley, A.J. Taylor, S.A. Wurm, John A. Z'graggen.
- No.40. WURM, S.A., ed. New Guinea area languages and language study, vol.3: Language, culture, society, and the modern world. 1977; lxxxvi+1449pp. (Hardbound: incl. 3 maps, 40 photographs), in two fascicles. ISBN 0 85883 131 7 and ISBN 0 85883 159 7
Articles by C. Abel, Henry L. Bell, Catherine H. Berndt, H. Myron Bromley, H.A. Brown, A. Capell, Lois Carrington, Emily Clarke, Anne M. Cochran, E.W. Deibler Jr, T.E. Dutton, Irënaus Eibl-Eibesfeldt, Franz-Josef Eilers, Joyce Franklin, Karl J. Franklin, Paul G. Freyberg, E. Fry, Alan Healey, L.R. Healey, R.K. Johnson, Adrienne Lang, Ranier Lang, Ralph S. Lawton, D.C. Laycock, Peter C. Lincoln, John Lynch, Howard P. McKaughan, Francis Mihalic, John Minogue, Peter Mühlhäusler, A.K. Neuendorf, Ebia Olewale, Andrew Pawley, G.L. Renck, Joan Rule, W.M. Rule, Gillian Sankoff, Robert P. Scott, Peter J. Silzer, A.J. Taylor, W.E. Tomasetti, Donald F. Tuzin, C.L. Voorhoeve, David Y.H. Wu, S.A. Wurm, John A. Z'graggen. Calvin Zinkel.
- No.41. FLIERL, W. and H. STRAUSS, eds Kâte dictionary. 1977; xxxv+499pp. (Hardbound: incl. 1 map). ISBN 0 85883 149 X
- No.42. NGUYỄN ĐĂNG LIÊM, ed. South-East Asian linguistic studies, vol.2. 1976; iv+262pp. (Hardbound: incl. 2 maps). ISBN 0 85883 144 9 and ISBN 0 85883 143 0
Articles by Paul K. Benedict, G. Diffloth, Eugénie J.A. Henderson, Judith M. Jacob, Philip N. Jenner, Joseph F. Kess, A. Llamzon, Ma. Teresita Martin, Malcolm Warren Mintz, Lili Rabel-Heymann, H.L. Shorto, Sidharta (Sie Ing Djiang), John U. Wolff.
- No.43. SMALLEY, William A., ed. Phonemes and orthography: language planning in ten minority languages of Thailand. 1976; xiii+347pp. (Hardbound) ISBN 0 85883 144 9
Articles by C.W. Callaway, Lois Callaway, Joseph R. Cooke, David Filbeck, David Hogan, E.R. Hope, J. Edwin Hudspith, Beulah M. Johnston, James A. Morris, Donald Schlatter, William A. Smalley, Peter Wyss.
- No.44. ZORC, David Paul The Bisayan dialects of the Philippines: subgrouping and reconstruction. 1977; xxiii+328pp. (Hardbound: incl. 9 maps). ISBN 0 85883 157 0
- No.45. NGUYỄN ĐĂNG LIÊM, ed. South-East Asian linguistic studies, vol.3. 1979; ix+326pp. (Hardbound) + 3 maps. ISBN 0 85883 144 9 and ISBN 0 85883 177 5
Articles by A. Capell, Soenjono Dardjowidjojo, Raleigh Ferrell, Jackson T. Gandour, Eugénie J.A. Henderson, Joseph F. Kess, Nguyễn Đăng Liêm, A. Kemp Pallesen, Alejandrino Q. Perez, Lawrence A. Reid, Alfonso O. Santiago, Patricia Stanley, Norman H. Zide, R. David Zorc.

Series C - Books (continued)

- No.46. HEALEY, Phyllis and Alan HEALEY *Telefol dictionary*. 1977; xix+358pp. (Hardbound). ISBN 0 85883 160 0
- No.47. PEREZ, A.Q., A.O. SANTIAGO and NGUYỄN ĐĂNG LIÊM, eds *Papers from the Conference on the Standardisation of Asian Languages, Manila, Philippines, December 16-21, 1974*. 1978; ix+386pp. (Hardbound: incl. 1 map). ISBN 0 85883 176 7
Articles by Abdullah Hassan, S. Takdir Alisjahbana, Asmah Haji Omar, Nelia G. Casambre, James R. Chamberlain, Sisir Kumar Das, Virgilio G. Enriquez, Andrew Gonzalez, Amran Halim, Astuti Hendrato-Darmosugito, Yahaya Ismail, Hans Kaehler, Harimurti Kridalaksana, Ernest W. Lee, Ferdinand E. Marcos, S.W. Rudjiati Muljadi, Nguyen Đang Liem, Fe T. Otones, Al Q. Perez, Ponciano B.P. Pineda, Wissanu Rawanking, Jack C. Richards, Alfonso O. Santiago, Takuji Sasaki, Bonifacio P. Sibayan, Hengtse Tu, Lars S. Vikör.
- No.48. GONZALEZ, Andrew *Pampangan: towards a meaning-based description*. 1981; xiii+402pp. (Hardbound) ISBN 0 85883 244 5
- No.49. NGUYỄN ĐĂNG LIÊM, ed. *South-East Asian linguistic studies, vol.4*. 1979; iv+436pp. (Hardbound). ISBN 0 85883 144 9 and ISBN 0 85883 201 1
Articles by Donald F. Barr, Maxwell Cobbey, James T. Collins, Joseph R. Cooke, Anthony Diller, Jack Gandour, John F. Hartmann, SamAng Hiranburana, Lou Hohulin, Mary Honts, Hope M. Hurlbut, Philip N. Jenner, Wilaiwan Kanittanan, Michael Kenstowicz, Joseph F. Kess, Anna Kwan-Terry, Nguyễn Đình-Hoà, Patcharin Peyasantiwong, Saveros Pou, Joan M. Rosen, Marmo Soemarmo, Tham Seong Chee, Michael R. Thomas, Udom Warotamasikhhadit.
- No.50. TRYON, D.T. *New Hebrides languages: an internal classification*. 1976; v+545pp. (Hardbound: incl. 7 maps). Reprinted 1979. ISBN 0 85883 152 X
- No.51. GLOVER, Warren W., Jessie R. GLOVER and Deu Bahadur GURUNG *Gurung-Nepali-English dictionary, with English-Gurung and Nepali-Gurung indexes*. 1977; xiii+316pp. (Hardbound). ISBN 0 85883 147 3
- No.52. MÜHLHÄUSLER, Peter *Growth and structure of the lexicon of New Guinea Pidgin*. 1979; xx+498pp. (Hardbound). ISBN 0 85883 191 0
- No.53. FRANKLIN, Karl J. and Joice FRANKLIN, assisted by Yapua KIRAPEASI *A Kewa dictionary, with supplementary grammatical and anthropological materials*. 1978; xi+514pp. (Hardbound: incl. 10 maps). ISBN 0 85883 182 1
- No.54. WURM, S.A., ed. *Australian linguistic studies*. 1979; xv+753pp. (Hardbound: incl. 18 maps, 7 illustrations). ISBN 0 85883 185 6
Articles by Barry J. Blake, A. Capell, Lois Carrington, Neil Chadwick, Jeffrey Heath, L.A. Hercus, Geoffrey N. O'Grady, Bruce Rigsby, M.C. Sharpe, Peter Sutton, Michael J. Walsh.
- No.55. LYNCH, John *Lenakel dictionary*. 1977; vii+167pp. ISBN 0 85883 165 1
- No.57. FOX, Charles E. *Arosi dictionary*. Revised edition with English-Arosi index prepared by Mary Craft. 1978; iv+598pp. (Hardbound: incl. 1 map). ISBN 0 85883 170 8
- No.58. THARP, J.A. and Y-Bhăm BUỒN-YĂ *A Rhade-English dictionary, with English-Rhade finderlist*. 1980; xi+271pp. (Hardbound). ISBN 0 85883 217 8
- No.59. BAUTISTA, Maria Lourdes S. *The Filipino bilingual's competence: a model based on an analysis of Tagalog-English code switching*. 1980; vi+386pp. (Hardbound). ISBN 0 85883 212 7

Series C - Books (continued)

- No.60. HEATH, Jeffery Basic materials in Mara: grammar, texts and dictionary. 1981; xiii+522pp. (Hardbound: incl. 1 map). ISBN 0 85883 247 X
- No.61. WURM, S.A. and Lois CARRINGTON, eds Second International Conference on Austronesian Linguistics: proceedings.
Fascicle one: 1978; xxii+1-688pp. (Hardbound: incl. 1 map, 2 photographs) ISBN 0 85883 184 8
Articles by Natalia Alieva, J.C. Anceaux, Robert Blust, Alice Cartier, Sandra Chung, Otto Chr. Dahl, Soenjono Dardjowidjojo, Isidore Dyen, Raleigh Ferrell, Jo-Ann Flora, Jeanne D. Gibson, A. Hakim Usman, R. Hardjadibrata, Hans Kähler, Joseph F. Kess, Don Laycock, Paul Jen-kuei Li, Paz Buenaventura Naylor, D.J. Prentice, Lawrence A. Reid, J.P. Sarumpaet, Ü. Sirk, H. Steinhauer, Claude Tchekhoff, Michael R. Thomas, John W.M. Verhaar, François Zacot, R. David Zorc.
Fascicle two: 1978; xxii-xxvii+688-1497pp. (Hardbound: incl. 6 maps) ISBN 0 85883 184 8
Articles by David G. Arms, Bruce G. Biggs, Ann Chowning, Ross Clark, Anne Cochran, Tom Dutton, Bryan Ezard, Jacques Bernard Michel Guy, S.P. Harrison, Marianne Haslev, Raymond L. Johnston, Yves Lemaître, P.C. Lincoln, John Lynch, Donald S. Marshall, Rodney F. Moag, Peter Mühlhäusler, Andrew Pawley, Albert J. Schütz, William J. Seiter, Andrew Taylor, Evelyn M. Todd, D.T. Tryon, D.S. Walsh, S.A. Wurm.
- No.62. SCOTT, Graham Fore dictionary. 1980; xiii+243pp. (Hardbound: incl. 1 map). ISBN 0 85883 226 7
- No.63. BROMLEY, H. Myron A grammar of Lower Grand Valley Dani. 1981; xiv+424pp. (Hardbound). ISBN 0 85883 223 2
- No.64. COPPELL, W.G. Austronesian and other languages of the Pacific and South-East Asia: an annotated catalogue of theses and dissertations. 1981; xiii+521pp. (Hardbound) ISBN 0 85883 238 0
- No.65. RANBY, Peter A Nanumea lexicon. 1980; xi+243pp. (Hardbound). ISBN 0 85883 227 5
- No.66. WURM, S.A. and Shirô HATTORI, eds Language atlas of the Pacific area, part 1: New Guinea area, Oceania, Australia. 1981; ii+72pp. (incl. 24 multicoloured maps, appropriate text materials, indexes). (Boxed set) ISBN 0 85883 239 9 and ISBN 0 85883 240 2 (Distributed by GeoCenter, Honigwiesenstrasse 25, D-7000 Stuttgart 80, Postfach 80 08 30, West Germany. Price DM 200.-)
- No.71. WORDICK, F.J.F. The Yindjibarndi language. 1982; xiv+390pp. (Hardbound: incl. 1 map). ISBN 0 85883 265 8
- No.73. FERRELL, Raleigh Paiwan dictionary. 1982; x+503pp. (Hardbound). ISBN 0 85883 264 X
- No.74. HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds Papers from the Third International Conference on Austronesian Linguistics. vol.1: Currents in Oceanic. 1982; vi+314pp. (incl. 8 maps). ISBN 0 85883 271 2
Articles by Tom Dutton, Jacques B.M. Guy, S.P. Harrison, Raymond L. Johnston, Don Laycock, John Lynch, Malcolm Ross, D.T. Tryon, D.S. Walsh.
- No.75. HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds Papers from the Third International Conference on Austronesian Linguistics. vol.2: Tracking the travellers. 1982; vii+331pp. (incl. 3 maps). ISBN 0 85883 275 5
Articles by J.C. Anceaux, James T. Collins, Isidore Dyen, Farid M. Onn, Mark Harvey, Hans Lapoliwa, Paul Jen-kuei Li, Teodore A. Llamzon, J. Noorduynd, Andrew K. Pawley, Lawrence A. Reid, Nicole Revel-Macdonald, Mangantar Simanjuntak, Stanley Starosta, C.L. Voorhoeve, John U. Wolff, Colin Yallop, R. David Zorc.

Series C - Books (continued)

IN PREPARATION:

- CAPELL, A. Futuna dictionary, with grammatical introduction.
- WURM, S.A. and Shirô HATTORI, eds Language atlas of the Pacific area. Part 2: Japan area, Philippines and Formosa, mainland and insular South-East Asia.
- LAYCOCK, D.C. Basic materials in Buin: grammar, texts and dictionary.
- STREICHER, J.-F. Jabêm-English dictionary.
- CAPELL, A. and H.H.J. COATE Comparative studies in Northern Kimberley languages, Australia.
- WURM, S.A., ed., with P. MÜHLHÄUSLER, D.C. LAYCOCK and T.E. DUTTON Handbook of New Guinea Pidgin.
- TRYON, D.T. and B.D. HACKMAN The languages of the Solomon Islands: an internal classification.
- WURM, S.A. John G. MEALUE and John Ini LAPLI Lö dai dictionary (Malo dialect), Northern Santa Cruz.
- HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds Papers from the Third International Conference on Austronesian Linguistics. vol.3.
- HALIM, Amran, Lois CARRINGTON and S.A. WURM, eds Papers from the Third International Conference on Austronesian Linguistics, vol.4.
- LYNCH J. ed., Studies in the languages of Erromango.
- BENDER, Byron W., ed., Studies in Micronesian linguistics (22 papers).

SERIES D - SPECIAL PUBLICATIONS

(BULLETINS, ARCHIVAL MATERIALS AND OTHER PUBLICATIONS)

- No. 1. Bulletin No.1. 1964; 9pp. ISBN 0 85883 072 8
Contributions by A. Capell, R.D. Eagleson, E.H. Flint, Susan Kaldor,
J. Lyons, S.A. Wurm.
- No. 2. Bulletin No.2. 1965; v+84pp. ISBN 0 85883 073 6
Contributions by J.R. Bernard, H. Bluhme, Christopher A.F. Court,
Robert R. Dyer, E.H. Flint, F.W. Harwood, Susan Kaldor, E.M. Liggins,
A. Murtonen, Anita Pincas, Hans Pollak, C.W. Ruhle, R.J. Zatorski.
- No. 3. WURM, S.A. New Guinea Highlands Pidgin: course materials.
1971; vii+175pp. ISBN 0 85883 074 4
- No. 4. WURM, S.A. Languages: Eastern, Western and Southern Highlands,
Territory of Papua & New Guinea. (Map in fourteen colours.)
1961. ISBN 0 85883 075 2
- No. 5. LAYCOCK, Don Materials in New Guinea Pidgin (Coastal and Lowlands).
1970; xxxvii+62pp. Reprinted 1974. ISBN 0 85883 076 0
- No. 6. NGUYEN ĐANG LIEM Four-syllable idiomatic expressions in Vietnamese.
1970; v+60pp. ISBN 0 85883 077 9
- No. 7. ELBERT, S.H. Three legends of Puluwat and a bit of talk. 1971;
ix+85pp. (incl. 1 map, 1 photograph). ISBN 0 85883 078 7
- No. 8. LANG, Adrienne, Katherine E.W. MATHER and Mary L. ROSE
Information storage and retrieval: a dictionary project. 1972; vii+151pp.
ISBN 0 85883 087 6
- No. 9. PACIFIC LINGUISTICS Index to Pacific Linguistics, Series A-D, as at
the end of 1970. 1971; iv+75pp. ISBN 0 85883 079 5
- No.10. PATON, W.F. Tales of Ambrym. 1971; xii+82pp. (incl. 1 map).
Reprinted 1978. ISBN 0 85883 080 9
- No.11. WURM, S.A., ed., with P. BRENNAN, R. BROWN, G. BUNN, K. FRANKLIN,
B. IRWIN, G. SCOTT, A. STUCKY, and other members of the Summer
Institute of Linguistics, New Guinea Branch Language maps of the
Highlands Provinces, Papua New Guinea. 1978; iii+21pp. (incl. 6 maps).
ISBN 0 85883 171 6
- No.12. DUTTON, T.E. Conversational New Guinea Pidgin. 1973; xviii+292pp.
Reprinted 1974, 1977, 1979, 1981. ISBN 0 85883 096 5
- No.13. GLOVER, Jessie R. and Deu Bahadur GURUNG Conversational Gurung.
1979; vii+216pp. ISBN 0 85883 192 9
- No.14. BARNETT, Gary L. Handbook for the collection of fish names in
Pacific languages. 1978; v+101pp. (incl. 1 map, 47 photographs,
3 drawings). ISBN 0 85883 175 9
- No.15. TRYON, D.T. & R. GÉLY, eds Gazetteer of New Hebrides place names /
Nomenclature des noms géographiques des Nouvelles-Hébrides.
1979; xxxiii+155pp. (incl. 1 map). ISBN 0 85883 194 5
- No.16. YOUNG, Maribelle Bwaidoka tales. 1979; viii+136pp. (incl. 1 map).
ISBN 0 85883 200 3
- No.17. STOKHOF, W.A.L., ed. ...with Lia SALEH-BRONKHORST Holle lists:
vocabularies in languages of Indonesia, vol.1: Introductory volume.
(Materials in languages of Indonesia, No.1: W.A.L. Stokhof, Series ed.)
1980; v+149pp. (incl. 2 maps). Reprinted 1981. ISBN 0 85883 213 5 and
ISBN 0 85883 214 3

Series D - Special Publications (continued)

- No.18. STANHOPE, John M. The language of the Rao people, Grengabu, Madang Province, Papua New Guinea. 1980; vii+28pp. (incl. 3 maps, 5 photographs). ISBN 0 85883 222 4
- No.19. STOKHOF, W.A.L. Woisika I: an ethnographic introduction. 1977: ix+74pp. (incl. 3 maps). Reprinted 1980. ISBN 0 85883 167 8
- No.20. CAPELL, A. and J. LAYARD Materials in Atchin, Malekula: grammar, vocabulary and texts. 1980; v+260pp. (incl. 1 map). ISBN 0 85883 231 3
- No.21. SCHEBECK, B. Texts on the social system of the At^{Yn}Yamaṭana people, with grammatical notes. 1974; xviii+278pp. + 1 photograph. ISBN 0 85883 102 3
- No.22. PATON, W.F. Customs of Ambrym (texts, songs, games and drawings). 1979; xv+98pp. (incl. 1 map, 4 photographs). ISBN 0 85883 189 9
- No.23. CLYNE, Michael, ed. Australia talks: essays on the sociology of Australian immigrant and aboriginal languages. 1976; viii+244pp. Reprinted 1978, 1980. ISBN 0 85883 148 1
Articles by M. Anne Bolitho, Michael Clyne, Robert D. Eagleson, R. McL. Harris, Ruth Johnston, Susan Kaldor, Manfred Klarberg, Stephen Muecke, Marta Rado, John Sandefur, Margeret C. Sharpe, J.J. Smolicz, Bruce A. Sommer, Brian A. Taylor, Elizabeth Thuan, Darrell T. Tryon.
- No.24. DUTTON, T.E. and C.L. VOORHOEVE Beginning Hiri Motu. 1974; xvii+259pp. Set of 6 cassettes (optional). Reprinted 1975. ISBN 0 85883 112 0
- No.25. Z'GRAGGEN, John A. The languages of the Madang District, Papua New Guinea. (Map) 1973. ISBN 0 85883 105 8
- No.26. LAYCOCK, D. Languages of the Sepik Region, Papua New Guinea. (Map) 1975. ISBN 0 85883 136 8
- No.27. WURM, S.A. Spreading of languages in the South-western Pacific. (Map) 1975. ISBN 0 85883 127 9
- No.28. STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONKHORST Holle lists: vocabularies in languages of Indonesia, vol.2: Sula and Bacan Islands, North Halmahera, South and East Halmahera. (Materials in languages of Indonesia, No.2: W.A.L. Stokhof, Series ed.) 1980; iv+325pp. ISBN 0 85883 213 5 and ISBN 0 85883 218 6
- No.29. DUTTON, Tom Queensland Canefields English of the late nineteenth century (a record of interview with two of the last surviving Kanakas in North Queensland, 1964). 1980; xiii+147pp. (incl. 3 maps, 2 photographs). ISBN 0 85883 224 0
- No.30. Z'GRAGGEN, J.A. A comparative word list of the Rai Coast languages, Madang Province, Papua New Guinea. 1980; xv+181pp. (incl. 1 map). ISBN 0 85883 232 1
- No.31. Z'GRAGGEN, J.A. A comparative word list of the Northern Adelbert Range languages, Madang Province, Papua New Guinea. 1980; xvi+178pp. (incl. 1 map). ISBN 0 85883 228 3
- No.32. Z'GRAGGEN, J.A. A comparative word list of the Mabusu languages, Madang Province, Papua New Guinea. 1980; xv+184pp. (incl. 1 map). ISBN 0 85883 233 X
- No.33. Z'GRAGGEN, J.A. A comparative word list of the Southern Adelbert Range languages, Madang Province, Papua New Guinea. 1980; xvi+97pp. (incl. 1 map). ISBN 0 85883 234 8

Series D - Special Publications (continued)

- No.34. LAPOLIWA, Hans A generative approach to the phonology of bahasa Indonesia. (Materials in languages of Indonesia, No.3: W.A.L. Stokhof, Series ed.) 1981; v+155pp. ISBN 0 85883 245 3
- No.35. STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONKHORST and Alma E. ALMANAR Holle lists: vocabularies in languages of Indonesia, vol.3/1: Southern Moluccas; Central Moluccas: Seram (I). (Materials in languages of Indonesia, No.4: W.A.L. Stokhof, Series ed.) 1981; iv+201pp. (incl. 1 map). ISBN 0 85883 213 5; 0 85883 242 9; 0 85883 243 7.
- No.36. HALIM, Amran Intonation in relation to syntax in Indonesian. (Materials in languages of Indonesia, No.5: W.A.L. Stokhof, Series ed.) 1981; vii+149pp. ISBN 0 85883 246 1
- No.37. NABABAN, P.W.J. A grammar of Toba-Batak. (Materials in languages of Indonesia, No.6: W.A.L. Stokhof, Series ed.) 1981; xxiv+146pp. (incl. 2 maps). ISBN 0 85883 254 2
- No.38. POEDJOSOEDARMO, Soepomo Javanese influence on Indonesian. (Materials in languages of Indonesia, No.7: W.A.L. Stokhof, Series ed.) 1982; viii+187pp. (incl. 1 map). ISBN 0 85883 261 5
- No.39. KARTOMIHARDJO, Soeseno Ethnography of communicative codes in East Java. (Materials in languages of Indonesia, No.8: W.A.L. Stokhof, Series ed.) 1981; xi+212pp. (incl. 1 map). ISBN 0 85883 255 0
- No.40. CARRINGTON, Lois and Miriam CURNOW Twenty years of Pacific Linguistics: an index of contributions to Pacific linguistic studies 1961-1981. 1981; vi+161pp. ISBN 0 85883 249 6
- No.41. STOKHOF, W.A.L. Woisika riddles. (Materials in languages of Indonesia, No.9: W.A.L. Stokhof, Series ed.) 1982; iii+74pp. (incl. 2 maps). ISBN 0 85883 257 7
- No.42. MCGREGOR, Donald E. and Aileen MCGREGOR Olo language materials. 1982; viii+155pp. ISBN 0 85883 262 3
- No.44. STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR Holle lists: vocabularies in languages of Indonesia, vol.3/2: Central Moluccas: Seram (II). (Materials in languages of Indonesia, No.10: W.A.L. Stokhof, Series ed.) 1981; iv+207pp. ISBN 0 85883 213 5; 0 85883 242 9; 0 85883 253 4
- No.45. SUHARNO, Ignatius A descriptive study of Javanese. (Materials in languages of Indonesia, No.11: W.A.L. Stokhof, Series ed.) 1982; xiv+175pp. ISBN 0 85883 258 5
- No.49. STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR Holle lists: vocabularies in languages of Indonesia, vol.3/3: Central Moluccas: Seram (III); Banda; Ambon (I). (Materials in languages of Indonesia No.15: W.A.L. Stokhof, Series ed.) 1982; vi+214pp. ISBN 0 85883 213 5; 0 85883 242 9; 0 85883 253 4
- No.50. STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR Holle lists: vocabularies in languages of Indonesia, vol.3/4: Central Moluccas: Ambon (II); Buru; Nusa Laut; Saparua. (Materials in languages of Indonesia No.16: W.A.L. Stokhof, Series ed.) 1982; iv+179pp. ISBN 0 85883 213 5; 0 85883 242 9; 0 85883 267 4

Series D - Special Publications (continued)

IN PREPARATION:

VERHEIJEN, J.A.J. Dictionary of Manggarai plant names.

VOORHOEVE, C.L., ed., The Makian languages, and their neighbours.
(Materials in languages of Indonesia, No.12: W.A.L. Stokhof, Series ed.)

COLLINS, James T. The historical relationships of the languages of
Central Maluku, Indonesia. (Materials in languages of Indonesia, No.13:
W.A.L. Stokhof, Series ed.)

TAMPUBOLON, Daulat Purnama Verbal affixations in Indonesian:
a semantic exploration. (Materials in languages of Indonesia, No.14:
W.A.L. Stokhof, Series ed.)

STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR
Holle lists: vocabularies in languages of Indonesia, vol.4:
Talaud and Sangir Islands. (Materials in languages of Indonesia No.17:
W.A.L. Stokhof, Series ed.)

STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR
Holle lists: vocabularies in languages of Indonesia, vol.5/1:
Indonesian New Guinea: Austronesian languages; Papuan languages, Digul area.
(Materials in languages of Indonesia No.18: W.A.L. Stokhof, Series ed.)

STOKHOF, W.A.L., ed., ...with Lia SALEH-BRONCKHORST and Alma E. ALMANAR
Holle lists: vocabularies in languages of Indonesia, vol.5/2:
Indonesian New Guinea: Northern languages; Central Highlands languages.
(Materials in languages of Indonesia No.19: W.A.L. Stokhof, Series ed.)

