

EASTERN MALAYO-POLYNESIAN: A SUBGROUPING ARGUMENT¹

ROBERT BLIJST

ABSTRACT

Essentially two positions have been taken regarding the subgrouping of the languages of extreme north-eastern Indonesia. Adriani and Kruijt (1914) maintained that there is a genetic unit which comprises the languages of southern Halmahera, certain languages of what was then called Geelvink Bay, and the languages spoken on the islands lying between these areas, including Misool. They rested their case on a small body of qualitative evidence, not all of which stands up to close scrutiny. Their analysis was nonetheless adopted by Esser (1938), who named this collection of languages the 'South Halmahera-West New Guinea group'. Dyen (1965), on the other hand, basing himself on the patterning of lexicostatistical percentages, partitioned these languages between at least two first-order Austronesian subgroups. Evidence is presented that five phonological innovations are in fact common to all SHWNG languages. Since the overall configuration of these changes is highly distinctive and is not known to be shared *in toto* with any other language, it seems simplest to follow Adriani and Kruijt in recognising a substantial subgroup of Austronesian languages in north-eastern Indonesia. This conclusion receives additional support from the discovery of a number of lexical items and irregular phonological changes that, so far as is known, are shared exclusively by SHWNG languages. Although Adriani and Kruijt did not comment on the wider subgrouping relations of the SHWNG languages, Dyen has suggested that some of these languages may belong to a lexicostatistically-defined subgroup that he calls the 'Moluccan Linkage'. Qualitative evidence is presented which appears to be explained more simply on the assumption that the SHWNG languages are the closest relatives of the great eastern subgroup of

Austronesian languages first recognised by Dempwolff (1937) and generally known today as 'Oceanic'. The term 'Eastern Austronesian' might be applied to this larger group, but in view of higher-level subgrouping conclusions reached recently in another work 'Eastern Malayo-Polynesian' is deemed more appropriate.

1. AUSTRONESIAN SUBGROUPING: SOME COMPETING VIEWS

All modern comparative work in Austronesian linguistics must make reference to Otto Dempwolff's fundamental *Vergleichende Lautlehre des austronesischen Wortschatzes* (1934-38), a contribution in which the comparative phonology of Austronesian languages was first placed in a completely systematic context. As is well known, the one outstanding subgrouping conclusion to emerge from Dempwolff's landmark publication concerned the existence of an eastern group of AN languages which included the languages of Polynesia together with many of the languages of Micronesia and Melanesia. Dempwolff called this putative subgroup 'melanesisch'.

In earlier publications Dempwolff implicitly excluded Chamorro (1920, 1924-25) and Palauan (1924-25) of Micronesia from his 'melanesisch' group. However, despite an isolated reference to Numfor (1924-25:318) as the westernmost language on the north coast of New Guinea which merges PAN *z/Z with *j, Dempwolff never precisely located the bundle of isoglosses separating 'melanesisch' from 'indonesisch' languages. In effect, then, although the criteria for the inclusion of a language in the 'melanesisch' group were fully explicit, the western boundary of the group was vaguely defined.

Partly as a result of Dempwolff's indeterminacy on this point subsequent researchers who accept his subgrouping conclusions have disagreed about the membership of the group in question. Grace (1955) proposed a division at "... (approximately) the boundary between Netherlands New Guinea and the Australian Trust Territory of New Guinea", and suggested 'Eastern Malayo-Polynesian' as an English equivalent of Dempwolff's 'melanesisch'. Milke (1958), on the other hand, maintained that the line of demarcation separating eastern from western Austronesian languages should be placed farther to the west so as to include all related languages as far as those of Sarera (formerly Geelvink) Bay in the eastern group. He further offered 'ozeanisch' as a substitute for Dempwolff's geographically prejudicial 'melanesisch'.

More recently Grace (1961, 1968, 1971) has agreed that all AN languages east of approximately 135 degrees east longitude except Chamorro and possibly Yapese appear to share an immediate common ancestor.

Following Milke he has adopted 'Oceanic' as the designation for this group. However, in contrast with Milke, Grace explicitly excludes Numfor-Biak from the Oceanic group on the grounds that this language maintains the distinction of PAN *b and *p, a contrast which he assumes to have been lost in Proto-Oceanic.

Some 20 years before the appearance of Dempwolff's *Vergleichende Lautlehre* the Dutch linguists Adriani and Kruijt (1914) published a massive three-volume study of the Bare'e-speaking Torajas of central Sulawesi. In the third volume of this important early contribution to the linguistics and ethnography of an Indonesian people the authors included a brief survey of the linguistic situation in north-eastern Indonesia. As part of this survey they appended (pp.302-5) a comparative wordlist of 101 items for the languages of east Makian, Buli, Sawai and Gane which clearly established the existence of a South Halmahera subgroup. But Adriani and Kruijt went beyond their initial claim and held that "... east Makian belongs with the languages of south Halmahera and the area of the Kalana Fat (Waigeo, Salawati, Misool), Numfor and its relatives". As evidence for this more inclusive subgroup they cited four features shared by the south Halmahera languages with Numfor:

1. final vowels are lost:

east Makian	Numfor	Malay	
ma-pin	bin	bini	'woman, wife'
kut	uk	kutu	'louse'
um	rum	rumah	'house'

2. apparently through stress shift many words developed a syncopated form:

east Makian	Numfor	Malay	
mtō	mga	mata	'eye'
plu	--	bulu	'body hair, feathers'
nhik (met.), Tagalog paniki			'large fruit bat'

3. the 3rd plural pronoun si is postposed as a pluraliser of nouns:

east Makian		Numfor	
mapin-si	'women'	mga-si	'eyes'
baba-si	'fathers'		
buk-si	'grandchildren'		

4. the 'reversed genitive' is used:

east Makian ²		
ai ni rage rage		'branch of a tree'
ai ni lu		'leaf of a tree'

The subgrouping value of these features will be discussed in a later section. For the present it is sufficient to mention that the same year that the third and concluding volume of Dempwolff's trilogy appeared in Germany also saw the appearance of the *Atlas van tropisch Nederland* in the neighbouring Netherlands. In this broad compilation of information on the largest of the Dutch tropical colonies the government linguist Esser presented a single-page classification of the languages of the then Netherlands East Indies. Esser recognised 17 groups of AN languages in the area at that time under Dutch political control. The two easternmost of these (nos. 16 and 17) are as follows:

16. South Halmahera-West New Guinea

1. South Halmahera languages
2. Numfor
3. Windesi
4. Kowiai
- etc.

17. Melanesian languages

1. Yautefa
2. Sarmi
- etc.

What is of particular interest to us here is that Esser followed Adriani and Kruijt in recognising a subgroup that consists of the languages of south Halmahera together with Numfor and its close relatives, and christened it the 'South Halmahera-West New Guinea' group.

In 1965 Isidore Dyen published a classification of the AN languages based on a computer tabulation of impressionistic cognate decisions for 352 basic vocabulary lists. In Dyen's classification the AN family is divided into 40 first-order subgroups, of which one (the 'Malayo-Polynesian Linkage') contains 129 members extending from Madagascar to Hawaii. Many of the remaining subgroups contain only a handful of languages, and some consist of a single representative. Surely the most striking feature of this result is the concentration of a large number of first-order subgroups in the western Solomons and the New Guinea-Bismarck Archipelago area. As Grace (1966) noted soon after the publication of Dyen's classification, this concentration of apparent first-order AN subgroups in western Melanesia conflicts with the evidence of comparative phonology, since many of Dyen's first-order subgroups are represented by members of Dempwolff's 'melanesisch' group.

Dyen also recognised five first-order AN subgroups in western New Guinea and the Moluccas. In that part of his classification which is relevant to the issues at hand he compared nine languages:

1. Buli (south-eastern Halmahera)
2. Minyafuin (Gebe island, west of Waigeo)
3. Biga (Wakde island, south of Waigeo)
4. As (north coast of the Vogelkop Peninsula)
5. Biak (Biak island, Sarera Bay)
6. Numfor (Numfor island, Sarera Bay)
7. Wandamen (west coast of Sarera Bay opposite Bintuni Gulf)
8. Yapen (Yapen island, Sarera Bay - language unspecified)
9. Waropen (east coast of Sarera Bay from the Mor islands in the south to beyond Kurudu island in the north)

Dyen's initial (i.e. computer-calculated) percentage for each pair of languages (indicated by the numbers 1-9 on either side of the colon) is given in Table 1:

TABLE 1

Cognate percentages on the Swadesh 200-word lexicostatistical test-list for the languages of south Halmahera and Sarera Bay, as reported by Dyen (1965)

1:2 (28.7)	5:6 (57.3)	5-6:9 (13.6)
1:3 --	5:7 --	7-8:9 --
1:4 --	5-6:8 (17.2)	
2:3 (23.4)	6:7 --	
2:4 --	7:8 (40.8)	
3:4 (23.1)		

The structure of Dyen's classification emerges from the recognition of varying orders of discreteness between collections of cognate percentages, and from the association of these orders of discreteness with corresponding levels of confidence that the statistical observations mirror the simple historical process of differentiation over time. To understand the procedure followed in constructing Dyen's tree for the AN family three terms are particularly important: critical percentage, basic percentage and critical difference.³ These terms are defined (Dyen 1965:18-19) as follows:

The percentage by which a language or group is classified together with other languages or groups is its critical percentage. A critical percentage which has been used to form a group is a basic percentage of that group The amount of difference between the lowest basic percentage of the group and the highest percentage of any member of the group with a non-member ... is called the critical difference.

The lexicostatistical lists are said to be adequate if they contain at least 174 equivalents of the 200 test-list meanings, and subadequate

if they do not. Based on the magnitude of their critical difference, subgroups are distinguished as belonging to one of five classes:

1. a subfamily, if the critical difference is 9.5 percentage points or greater
2. a genus, if the critical difference is 8.0-9.4
3. a cluster, if the critical difference is 5.0-7.0
4. a hession, if the critical difference is 2.5-4.9
5. a linkage, if the critical difference is less than 2.5

There is no explicit indication as to how a critical difference between 7.0 and 8.0 would be interpreted in terms of the hierarchy of discreteness proposed, though as will be seen below Dyen interprets percentages in this range as belonging to the cluster class.

Table 1 can now be interpreted as follows. Buli and Minyafuin score 28.7% on the lists available for comparison (which are said to be adequate). Since this is the highest percentage shared by either language with another language in Dyen's sample it becomes the critical percentage for uniting Buli and Minyafuin in a (two-member) subgroup. At the same time this figure is the basic percentage used in determining the level of discreteness of the subgroup represented by these two languages. The next highest percentage shared by either Buli or Minyafuin with a third language is 23.4, shared by Minyafuin with Biga. Since the critical difference separating Buli and Minyafuin on the one hand from Biga on the other falls between 5.0 and 7.0 ($28.7 - 23.4 = 5.3$), the subgroup consisting of Buli and Minyafuin is called a cluster (the 'Bulic Cluster').

It will be noticed immediately that cognate percentages are not reported for every pair of languages (hence the dashes following Buli:Biga, Buli:As, etc. and the lack of consideration of the percentages holding between languages 1-4 *vis-à-vis* 5-9). This practice is followed because given Dyen's assumptions only critical differences are relevant to determining the structure of the classification. As observed above, Dyen defines the critical difference as the difference between the lowest basic percentage of a group and the highest percentage with a non-member. Since 28.7 is the only basic percentage for the Bulic Cluster and since Minyafuin scores somewhat higher than Buli with Biga, the Buli-Biga percentage does not figure in determining the critical difference for the Buli-Minyafuin subgroup, and therefore is not reported. Similarly, once the Buli-Minyafuin subgroup is established as a cluster by its critical difference with Biga there is no need to report its percentages with As, since As also scores higher with Biga than with any other language.

As can be seen, the percentage shared by Biga with As is not significantly different from that shared by Biga with the Bulic Cluster. For this reason Biga and As are combined with the Bulic Cluster as coordinate members of a somewhat larger genetic unit. The next highest percentage shared by any of these languages with another language is 15.8, between Ivatan of the Batanes islands north of Luzon in the Philippines and an unspecified member of the Bulic Cluster. Although in this case the critical difference is over 7.0 ($23.1 - 15.8 = 7.3$), Dyen assigns cluster status to the resulting four-member subgroup (the 'Bigic Cluster').

A qualification should be introduced at this point. Dyen classifies all groups having at least a genus-level order of discreteness as closed groups, and others as open groups. Given the relatively low level of confidence that follows from their statistical definition, and assuming that sampling errors have not substantially affected the results with closed groups, it is clear that open groups have the highest probability of appearing to be discrete by chance. The method of determining critical differences that has been described is in fact restricted to open groups. For groups that have the lowest probability of appearing to be discrete by chance (i.e. subfamilies and genera) the percentages of members with non-members were averaged. In Dyen's words (p.19) "These averaged percentages are regarded as the percentages of the closed group. In effect members of closed groups are treated like dialects of the same language or lists from the same language."

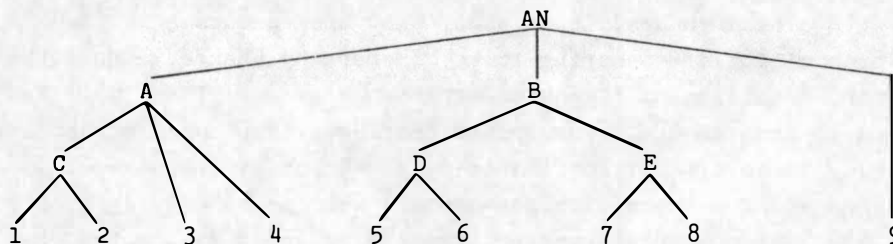
In accordance with this principle Biak and Numfor are united in the Biakic Subfamily (critical difference $57.3 - 17.2 = 40.1$), and Wandamen and Yapen are united in the Wandamic Subfamily ($40.8 - 17.2 = 23.6$) through the use of a basic percentage (17.2) which is the average of the two unreported percentages for Biak-Yapen and Numfor-Yapen (hence the conjunction of the two Biakic languages as 5-6:8 in Table 1). The highest percentage holding between a member of either the Biakic Subfamily or the Wandamic Subfamily and another language is 14.6, shared by Kuiwai of the south-western coast of New Guinea with the Biakic Subfamily (i.e. with the average of the two unreported percentages Kuiwai:Biak and Kuiwai:Numfor). Given the relatively small critical difference separating this collection of four languages from Kuiwai ($17.2 - 14.6 = 2.6$), Dyen labelled the putative subgroup composed of the Biakic and Wandamic Subfamilies a hesion (the 'Geelvink Hesion').

The highest percentage shared by Waropen with another language was found to be 13.6 with the (averaged) Biakic Subfamily. Waropen accordingly was regarded as an isolate, and assigned to a category of 'ungrouped' languages (pp.27-8).

From the above description it can be seen that the nine languages listed in Table 1 fall into three subgroups on the highest level of inclusion (the Bigic Cluster, the Geelvink Hesion, Waropen). Stated somewhat differently, in Dyen's classification none of these lexico-statistically-defined subgroups is hierarchically included in a larger subgroup short of Austronesian itself: each is a first-order subgroup of the Austronesian Linkage. This result is diagrammed in Table 2:

TABLE 2

Tree-diagram representation of Dyen's (1965) initial classification of the AN languages of south Halmahera, the northern Vogelkop Peninsula and Sarera (= Geelvink) Bay



A = Bigic Cluster, B = Geelvink Hesion, C = Bulic Cluster, D = Biakic Subfamily, E = Wandamic Subfamily

1 = Buli, 2 = Minyafuin, 3 = Biga, 4 = As, 5 = Biak, 6 = Numfor, 7 = Wandamen, 8 = Yapen, 9 = Waropen

In an appendix which was added after the structure of his classification had already been determined, Dyen noted (p.59) that significant errors in the percentages for certain language pairs on the north coast of New Guinea had been pointed out to him in personal correspondence by George Grace. Since the errors were of such a surprisingly high order of magnitude (e.g. Yotafa:Tobat1 33.8 (original): 65.4 (revised), Yotafa:Kayupulau 20.8:59.7, etc.) a restricted restudy was made for other languages in this general area, evidently giving greater attention to the synchronic morphology and comparative phonology than was given when the original cognate decisions were reached. As a result similar changes were also made in the percentages *inter se* for members of the Bigic Cluster and Geelvink Hesion, as shown in Table 3 (original percentage precedes and revised percentage follows the slash):

TABLE 3

Revised cognate percentages for languages originally assigned to the Bigic Cluster and to the Geelvink Hesion

1:2 (28.7/45.5)	5:6 (57.3/58.7)	5-6:9 (13.6)/
1:3 --	5:7 --	5-8:9 (27.7)
1:4 --	5-6:8 (17.2/31.7)	
2:3 (23.4/43.5)	6:7 --	
2:4 --	7:8 (40.8/60.6)	
3:4 (23.1/44.2)		

Because the upward revision of the internal percentages for members of the original Bigic Cluster was not accompanied by a corresponding upward revision of percentages with non-members (the recalculation actually lowered the percentage computed with Ivatan from 15.8 to 14.2) the order of discreteness of the Bigic Cluster was increased to the subfamily level. At the same time the closer sequencing of the revised percentages within the Bigic Subfamily obliterated the earlier evidence for a Bulic Cluster.

Although the internal percentages for members of the Geelvink Hesion were also revised upward, this revision was accompanied by a corresponding upward revision in the percentages scored by these languages with Waropen (13.6 to 27.7). As a result the revised basic percentage of the Geelvink group came to exceed the revised percentage shared with Waropen by only four points (31.7 less 27.7), thus leading to no change in the original nomenclature.

It was seen above that all three major subgroups represented in Table 2 were regarded as primary branches of the Austronesian Linkage (or Family). Given the rather extensive revision in some of the re-computed percentages for these groups one might expect some significant structural changes to result in the classification. Dyen maintains (p.58) that "There is some reason to believe that this is the case for the Sarmic Subfamily and the Geelvink Hesion whereas it is perhaps not true for the Bigic Subfamily."

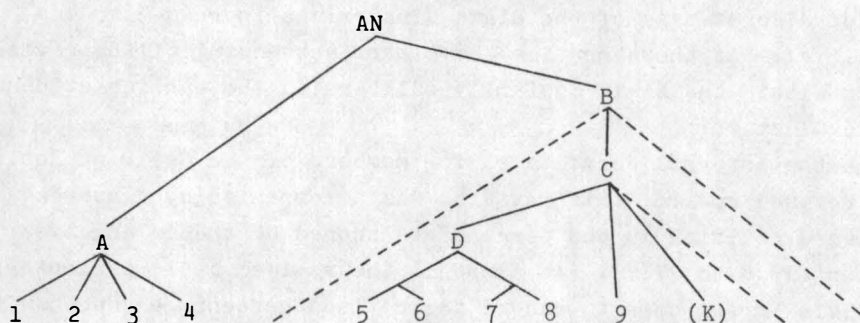
This difference in the structural consequences of the revised percentages results from the discovery of intermediate percentages linking the Geelvink Hesion and Waropen to larger collections of languages short of the Austronesian Linkage itself, while no such intermediate percentages were found in the case of the Bigic Cluster. Thus, Dyen notes that not only does the Geelvink Subfamily average 27.7 with Waropen, but Biak shows 28.2 with Kuiwai of the Moluccan Linkage (one of seven first-order subgroups of the Malayo-Polynesian Linkage, extending from Flores in the west to the Bomberai Peninsula of New

Guinea in the east).⁴ Based on these observations he concludes (p.59) "All of this appears to suggest a closer relationship of the Sarmic and Geelvink groups as well as Waropen with the Moluccan Linkage than has thus far come to light."

In summary, then, the final published statement of Dyen's position is that the nine languages in question fall into two first-order AN subgroups: 1. the Bigic Subfamily, 2. the Malayo-Polynesian Linkage. All of those languages that belong to the Malayo-Polynesian Linkage are also members of one of its primary subgroups, the Moluccan Linkage. These relationships appear in Table 4:

TABLE 4

Tree-diagram representation of Dyen's (1965) revised classification of the AN languages of south Halmahera, the northern Vogelkop Peninsula and Sarera Bay



A = Bigic Subfamily, B = Malayo-Polynesian Linkage, C = Moluccan Linkage, D = Geelvink Hesion

1 = Buli, 2 = Minyafuin, 3 = Biga, 4 = As, 5 = Biak, 6 = Numfor, 7 = Wandamen, 8 = Yapen, 9 = Waropen, (K) = Kuiwai

Broken lines indicate a non-specific number of additional branches represented by lists used in Dyen's classification

In contrast to this picture of relatively great linguistic diversity in north-eastern Indonesia I will propose in the following pages

- 1) that strong qualitative evidence can be adduced in support of a South Halmahera-West New Guinea (SHWNG) subgroup;
- 2) that the SHWNG subgroup must encompass all AN languages of Halmahera and Sarera Bay together with the intervening islands, including Misool;
- 3) that the SHWNG and Oceanic subgroups form co-ordinate branches of a larger eastern subgroup of the AN family for which Grace's (1955) term 'Eastern Malayo-Polynesian' can appropriately be revived as a designation.

2. QUALITATIVE EVIDENCE FOR A SOUTH HALMAHERA-WEST NEW GUINEA SUBGROUP

As it has not been possible to collect material for the SHWNG languages in the field, the material in this section will be drawn primarily from languages for which published dictionaries or extensive wordlists are available. For the languages of south Halmahera the principal witness will thus be BULI as recorded in the grammar and wordlist of Maan (1940, 1951). For the languages of Sarera Bay the principal witness will be NUMFOR as recorded in the dictionary of van Hasselt and van Hasselt (1947), with occasional remarks on BIAK as recorded by Suparno (1975).⁵ Other references will be given in a later subsection, at which point all available materials for SHWNG languages will be surveyed.

2.1. PHONOLOGICAL EVIDENCE

Even a superficial lexical comparison of Buli and Numfor is sufficient to establish that these languages have participated in several strikingly similar phonological developments, as was first pointed out in part by Adriani and Kruijt in 1914. More intensive comparison reveals that Buli and Numfor have undergone the following phonological shifts (i.e. phonetic changes not leading to phonemic merger), and mergers. Only those changes that are shared in common are numbered. Rules that are described by name (e.g. Apocope 1) are marked by a plus if there is evidence of their application, and by a minus if there is not. In all other cases a minus (= hyphen) is used to indicate that a description of the reflex in question is deferred to a later point so that similar reflexes can be grouped together in each language. Thus *-d- > Buli r (rare), Numfor minus (-) means that the Numfor reflex of *-d- (=r) is by-passed at this point so that it can later be described together with other sources of r. Subscripts mark the number of syllable peaks from the end of a morpheme (hence *e₂ occurs in the penult, but *e₁ in the ultima).

PAN	BULI	NUMFOR
	Consonant shifts	
1. *p	f	f
*C, t, T	c (rare)	k
*b	p	(b)
*-d-	r (rare)	-
	Vowel shifts	
2. *e ₂	o	o
	Epenthesis	
3. *a-	ya	ya
	Consonant mergers	
4. *C, t, T/ _i	s	s
*c	s	s?
*-j-	s	s
*s	s	s
*z, Z	-	s?
5. *k	∅	∅
*q, ?	∅	∅
*H, S, x	∅	∅
*-j	-	∅
*R	∅	r
*-ŋ	∅ (rare)	-
6. *d, D	l	r
*z, Z	l	r?
*l	l	r
*r	l	r
*-j	l	-
7. *n	n	n
*ñ	n	n
*ŋ	ŋ	n
*-n	ŋ (rare)	-
8. Cluster reduction (C merges with zero)	+	+
	Vowel and diphthong mergers	
9. *a	a	e
*e ₁	a	e
10. *i	i	e
*u ₂	i	e
*a ₁	i (rare)	e
11. Apocope 1 (V merges with zero)	+	+
12. Syncope (V merges with zero)	+	+

PAN	BULI	NUMFOR
13. Apocope 2 (original final vowels and final vowels from diph- things merge with zero)	+	+
Contraction	-	+

Evidence for the numbered changes (i.e. those shared in common) is as follows:

1. *p > B, Nf f: *pitu > B fit, Nf fik '7'
2. *e₂ > B, Nf o: *depa > B lof, Nf rof 'fathom'
3. *a- > B, Nf ya-: *añam > B yanam, Nf yanem 'plait'
4. *C, t, T/ _i, *c, *-j-, *s > B, Nf s: *bi(t)il > B bi-bisil, Nf biser 'hunger, hungry', *cencen > B sosan 'dense, stopped up',⁶ *ñajan > B ñasan 'name', Nf nasan '(personal) title', *susu > B, Nf sus 'breast'
5. *k, *q, *ʔ, *H, *S, *x > B, Nf Ø: *kawil > B awil, Nf awir 'fish-hook', *qatep > B yataf 'thatch, roof', *qanjasaw > Nf yas 'day', *Sajek > Nf yas 'sniff, kiss',⁷ *ʔenem > B wonam, Nf (w)onem '6', *baHi > B ma-pñ, Nf bin 'woman, wife, female', bai 'woman, wife (respectful term of address)', *Sun(e)qap > B un-unaf, Nf unef 'fish scale', *x₂epat > B fat, Nf fiak '4'
6. *d, *D, *z, *Z, *l, *r > B l, Nf r: *depa > B lof, Nf rof 'fathom', *kuDen > B ulan, Nf uren 'clay cooking-pot', *tazim > B dalim 'sharp', *quZan > B ulan 'rain', *ZuRuq > Nf rur 'sap, juice, gravy',⁸ *lima > B lim, Nf rim '5', *ra(q)uŋ > B lau 'to howl, of dogs', *buri > Nf pur⁹ 'after'
7. *n, *ñ > B, Nf n: *nanaq > B, Nf nan 'pus', *qañam > B yanam, Nf yanem 'plait'
8. Cluster reduction: *DemDem > B lolam 'brood, meditate', *kalibənən > B aibobaŋ, Nf apop 'butterfly'
9. *a, *e₁ > B a, Nf e: *pa(n)pan > B fapan, Nf am-bafen 'plank', *ʔenem > B wonam, Nf (w)onem '6'
10. *i, *u₁, *a₁ > B i, Nf e: *laŋiC > B laŋit 'sky', Nf nanek 'sky, heaven, the highest power (word by which oaths are sworn)', *Rusuk > B usi 'ribs', *lumut > Nf rumek 'moss, algae', *ʔuRaC > B uit, Nf urek 'vein, vessel; tendon'

11. Apocope 1: *SabaRat 'north-west monsoon' > B pat 'west; west wind'.
Nf barek 'west', wam barek 'west wind, west monsoon'
12. Syncope: *maCa > B mta, Nf mga 'eye'
13. Apocope 2: *banua 'inhabited area' > B pnu, Nf menu 'village',
*tebuS(u) > B top, Nf kob 'sugarcane', *bunuq > B pun, Nf mun
'kill', *laŋaw > B laŋ 'a fly', Nf ran 'k.o. small fly'

Problems of analysis

The initial phonological evidence for a South Halmahera-West New Guinea subgroup has now been presented in a form that permits ready comparison. One can hardly fail to be struck by the fact that the phonological development of Buli and Numfor is virtually identical in its major outlines, differing mainly in details that can be attributed to relatively recent changes in the individual histories of the two languages (viz. all unnumbered changes, as *C,t,T > B c (rare), Nf k, *b > B p, etc.). It is certainly remarkable that no less than three shifts and 10 mergers are shared between the two languages. Given the number of phonological innovations identified this result normally would be indicative of a subgrouping relationship even if the changes in question were of a type widely distributed in other AN languages. Yet several of these changes taken individually are highly distinctive (e.g. 2, 6, 9, 12) and others are of a type that, though found in some other languages, is not so common as to deprive them individually of subgrouping value (e.g. 4, 5, 11, 13).

One, however, does not ordinarily evaluate subgrouping arguments on the basis of individual innovations, but rather on the basis of collections of innovations, and here the grounds for regarding Buli and Numfor as subgroup relatives seem particularly strong. Thus the 13 common innovations noted above form a highly distinctive pattern of change, a part of which - as we shall see - clearly marks off members of the SHWNG group from all other AN languages.

Perhaps even more impressive, the ordering of the changes in the two languages is largely identical. To cite one of several possible examples, both Buli and Numfor have lost *k, and added a palatal glide before a low vowel that was originally initial or that had come to be initial as a result of the loss of an initial laryngeal. But in both languages words that originally began with *ka- now begin with a, thus requiring the assumption that glide epenthesis preceded loss of initial *k.

If it should prove possible to take all or even most of the foregoing observations at face value there can be little question that it

will be difficult to explain them without the assumption that Buli and Numfor share an immediate common ancestor. In fact, given the depth of most historical analyses in comparative Austronesian studies to date (which ignore ordering requirements altogether), it can be said that this assumption is almost unavoidable. However, the two principal analytical/theoretical problems that must be faced by the advocate of any subgrouping claim based on shared phonological innovations are those of rule borrowing (= wave-like spread of phonological changes) and drift (= parallel evolution). The evidence that shared changes are in fact independent developments can be direct or indirect. We will consider the indirect evidence first.

Waves and drift: the indirect evidence

Two of the innovations that Buli and Numfor share are the merger of *k with *q, *ʔ, *H, *S and *x as zero (no.5), and loss of final vowels (no.13). The loss of *k is seen e.g. in

PAN	Buli	Numfor	English
*kaSiw	ai	a(i)	'wood, tree'
*iS(e)kan	ian	in	'fish'
*manuk	mani	man	'bird; fowl'

and the loss of final vowels in

*depa	lof	rof	'fathom'
*i-a	i	i	'3rd sg. actor'
*telu	tol	kior	'three'

Both changes appear in *kuCu > B ut, Nf uk 'head louse'. Moreover, it is clear from examples such as

*babaq	pap	bab	'beneath, under'
*bunuq	pun	mun	'kill'

that the loss of final *q preceded loss of final vowels in both languages (or else the last-syllable vowel in forms that originally ended in *q presumably would be retained).

What is noteworthy in the present connection is that when a word ended in *k the vowel that preceded it has been retained in Buli, but was lost in Numfor:

*ma-esak	masa	--	'cooked, ripe'
*manuk	mani	man	'bird; fowl'
*Rebek	opa	rob	'to fly'
*Rusuk	usi	--	'rib'
*Sajek	--	yas	'sniff, kiss'
*(t)uZuk	culi	--	'point out, indicate'

final vowels are not permitted after a consonant) would at first have been identical in the two linguistic traditions. The loss of final *R in Buli might then have introduced some final vowels, thereby breaking this constraint prior to the loss of final *k. When final *k was lost the vowels that were thus made final would remain in Buli since they violated no morpheme structure constraint, but would disappear in Numfor, since they did violate such a constraint. However, such a hypothesis fails to explain why the loss of final *R in Buli would permit the violation of an earlier morpheme structure constraint, while the loss of final *k in Numfor would not. All in all, then, the safest course probably is to adopt the rule-borrowing solution. As will be seen, the broader comparative picture lends further support to this conclusion.

Waves and drift: the direct evidence

As we have just seen, there is indirect evidence in the phonological histories of Buli and Numfor that some of the changes shared by these languages are not due to inheritance from an immediate common ancestor in which they occurred. It is thus a matter of some interest to note that comparative evidence from closely related languages also suggests that these changes have occurred independently in the two languages. To fully appreciate this evidence it will be necessary to collect all available information on the phonological history of other AN languages in Halmahera and the general region of Sarera Bay. In this expansion of our inquiry it will be possible not only to bring comparative evidence to bear on the problem of distinguishing shared innovations due to diffusion and drift from those probably due to innovation in an immediate common ancestor, but also to investigate - however briefly - the scope of any resulting subgroup.

THE SOUTH HALMAHERA LANGUAGE GROUP

The subgrouping connection of the languages of south Halmahera is so close as to be apparent on inspection. This indeed has been the view of all past investigators (de Clercq 1890, Adriani and Kruijt 1914, Esser 1938, introductory remarks in Maan 1940, 1951,¹¹ Dyen 1965, in so far as his 'Bulic Cluster' can be regarded as encompassing the south Halmahera languages as a group, Masinambow 1972), and a convincing demonstration of this thesis based on exclusively shared lexical innovations was given as early as 1914 by Adriani and Kruijt (pp.302-5).

In addition to Maan's wordlist and grammar of Buli, which is by far the fullest description yet available for any south Halmahera language, a much earlier sketch of Buli is given by Adriani and Kruijt (1914:309-35),

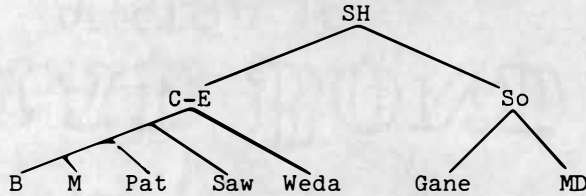
who also include thumbnail sketches of Maba (335-8), Patani (338-41), Sawai (341-4), Weda (344-5) and Gane (345-7). Further published data are also available for Gane (van der Crab 1862, Wallace 1869; the latter being the principal source for Adriani and Kruijt, but the former, much more extensive vocabulary not mentioned) and east Makian (also known as Makian Dalam; van der Crab 1862, de Clercq 1890). Wherever possible the Gane material has been supplemented through unpublished fieldnotes collected by Dik Teljeur of the Department of Social and Cultural Anthropology, Free University of Amsterdam, during the course of anthropological fieldwork in the Gane-speaking region of extreme southern Halmahera. In addition the writer has had access to a Swadesh 200-item lexicostatistical test list collected for Maba, and part of the same list collected for Makian Dalam.¹²

Reflexes of over 180 PAN reconstructed forms have been identified in the survey materials culled from the above sources, and these permit the following conclusions:

- 1) all of the languages of south Halmahera (including Makian Dalam and Kayoa) have undergone changes 1, 2, 3, 4, 6, 7, 8, 9, 10,¹³ 11, 12 and 13
- 2) it is clear from the materials published by van der Crab (1862) and Wallace (1869) that all of the languages of Misool for which information is available have also undergone these changes.

It might be said by way of preface that even from the limited materials now available it is clear that Maba is very closely related to Buli, perhaps being a dialect of the same language. A similar view is expressed by Adriani and Kruijt (1914), and by Maan (1951:5). The thumbnail sketches that Adriani and Kruijt give for Patani, Sawai and Weda suggest further that the whole of the central and eastern portions of southern Halmahera forms a dialect continuum in which even the extremes do not differ greatly. The southernmost languages, however, must be excepted from this statement.

Gane is spoken over the lower half of the southern peninsula of Halmahera, and is said by van der Crab (1862) to be spoken also on the island of Kayoa. The relationship between Gane and Makian Dalam (interior of Makian island) appears to be about as close as that between Buli and Maba - i.e. the two regions may represent only slightly differing areas in a dialect continuum. A major bundle of phonological and lexical isoglosses, however, separates this southern dialect group from the central and eastern dialect group. The subgrouping picture for the south Halmahera languages can thus be diagrammed approximately as follows:



SH = south Halmahera, C-E = central-eastern, So = southern,
 B = Buli, M = Maba, Pat = Patani, Saw = Sawai, MD = Makian
 Dalam

Limitations of space prevent my stating the full evidence for conclusions 1 and 2, or my justifying all nodes in the above family tree at the present time. Accordingly in this summary I will attempt only to justify the top node in the south Halmahera family tree (i.e. the existence of a south Halmahera subgroup), the southern node (i.e. the existence of a southern south Halmahera dialect cluster), and the inclusion of the languages of Misool in the south Halmahera group. We will begin with the southern node.

The hypothesis that Gane and Makian Dalam form a dialect area apart from all other languages of Halmahera can be justified by the existence of a substantial number of lexical isoglosses that appear to be shared exclusively by the southern languages. No attempt will be made to state all relevant evidence that can be extracted from the available materials. The following examples, however, should be sufficient to establish the point.

TABLE 5

Evidence for a southern dialect area among the south Halmahera languages. The following lexical items appear to be shared by Gane and Makian Dalam apart from all other AN languages.

Gane	Makian Dalam	English
1. lo-poas	poas	'voyage'
2. kuyu-t	kuyo	'fingernail'
3. poyo-d	poyo	'head'
4. graak	garak	'take fright'
5. betol	batol	'star'
6. wilo-t	wilo	'lip'
7. nan halalm	nan halain	'midday'
8. mauka	mauka	'wall'
9. hatatal	hatetal	'war'
10. kawiji	kawiji	'porridge'
11. necapa	encepa	'plume'
12. im	im	'talk'
13. bulo	bulo	'raw'

Gane	Makian Dalam	English
14. lisko	likso	'edge'
15. awoyan	wayan	'right (side)'
16. hawol	hawol	'to call'
17. glil	glil	'cross-eyed'
18. non	nan	'sharp'
19. mla	na-mna	'swift'
20. lai	la-lai	'stone'
21. toban	taban	'time'
22. dobadoba	dabadaba	'garden'
23. hatut	hatut	'quarrel'
24. litlit	litlit	'owl'
25. sobal	sobal	'voyage'
26. liko	liko-t	'skin, hide'
27. atah	eta	'to find'
28. kiu	kiu	'fear, afraid'
29. lomog	lomog	'friend'
30. ŋan sosopo	ŋan sasopo	'west'
31. mei	me	'who?'
32. liklik	lik	'behind'
33. hadumito-do	hadomit	'refuse, decline'
34. maleo	maleo	'trade'
35. saŋu	saŋu	'other'
36. loni manitap	odoni manitap	'to answer'
37. tanawan	tonawan	'to love'
38. lipan	lipan	'bury'
39. lainpe	laipe	'begin'
40. kutu	kutu	'small'
41. hintun	haitun	'give birth'

On the assumption that the major split within the south Halmahera group (as diagrammed above) can now be taken as established, it is possible to justify claim (1) by reference only to the southern languages, since any feature of the phonological history of Buli which is shared with them presumably will prove to be shared with all other SH languages. Again, limitations of space make it impossible for me to present all relevant evidence that has been collected to date. Representative support for claim (1) is as follows (B = Buli, G = Gane, MD = Makian Dalam):

1. *p > B f, G f ~ h, MD h: *pa(n)iki > G fnik, MD nhik, nihik (met.)
'flying fox', *panaw > G, MD han 'go, walk', *pia > G fia ~ hia,
MD hia 'good'
2. *e₂ > B, G, MD o: *denger > G loŋa, MD ha-loŋa 'hear', *depa > G
lof, MD loh 'fathom', *telu > G, MD tol 'three'
3. *a- > B, G, MD ya-: *aku > G, MD yak '1st sg. actor', *asin > G
m-yasiŋ, MD yasi 'salty', *qapuR > G yafi ~ yahi, MD yahi 'lime,
chalk'
4. *C,t,T/_i, *c,*-j-, *s > B, G, MD s:¹⁴ *pija > G fis 'how much, how
many?', *esa > G, MD -so 'one', *(s)aRu > G, MD sei 'comb'
6. *d,D, *z,Z, *l, *r > B, G, MD l: *denger > G loŋa, MD ha-loŋa 'hear',
*DuSa > G, MD -lu 'two', *Zalan > G lolan, MD lolan 'path, way,
road', *lima > G, MD -lim 'five'
7. *n, *ñ > B, G, MD n: *niuR > G, MD niwi 'coconut tree', *ñamuR >
MD nomi 'dew'
8. Cluster reduction is found in Buli and Gane; data for Makian Dalam
are lacking: *bitbit 'grasp with the fingers' > G a-pipit 'pinch',
*bunbun-an > G puŋan 'ridgepole, ridge of the roof', *busus busus
'leak, spill, pour through' > G ta-bubus 'diarrhoea'
9. *a, *e₁ > B a, G, MD a, o: *bunbun-an > G puŋan 'ridgepole, ridge
of the roof', *esa > G, MD -so 'one', *kuDen > G kulan, MD kulan
'earthen cooking-pot', *lipen > G ha-hlo-d (met.), MD la-lho 'tooth'
10. *i, *u₂ > B, G, MD i: *manuk > G, MD manik 'bird; fowl', *qapuR >
G yafi ~ yahi, MD yahi 'lime, chalk'
11. Apocope 1 is found in all three languages: *qaCeluR > G tolo ~
toli, MD tolo 'egg', *SabaRat 'north-west monsoon' > G poat 'west,
west wind'
12. Syncope is found in all three languages: *maCa > G, MD mto
*pa(n)iki > G fnik, MD nhik, nihik (met.) 'flying fox'
13. Apocope 2 is found in all three languages: *telu > G, MD tol
'three', *salaq > G en-col, MD in-cal 'wrong, in error',
*be(R)(s)ay > G poas 'canoe paddle', *panaw > G, MD han 'go, walk'

With some partial indeterminacies due to gaps in the data (e.g. earlier *t > s/_i), then, 12 of the 13 changes shared by Numfor with Buli are also shared with Gane and Makian Dalam. As can be seen from the examples cited, Gane and Makian Dalam also share some changes with Buli apart from Numfor (as *b > p and *R > Ø). On the basis of this

evidence the existence of a south Halmahera language group can be taken as established.

The classification of Misool¹⁵ as a south Halmahera language is based on a combination of phonological and lexical innovations. Since the lexical innovations are more specific than the phonological innovations (pointing clearly to a subgrouping relationship with the south Halmahera languages rather than to simple inclusion in the SHWNG group), a representative sample of these is given in Table 6:

TABLE 6

Evidence for the inclusion of Misool in the south Halmahera language group. The following lexical items appear to be exclusively shared by Misool with various languages of southern Halmahera. For convenience the languages of southern Halmahera unless otherwise noted are represented by Buli

Misool	Buli	English
1. te-	ti- (Maba)	verb prefix ¹⁶
2. owon to	to	'ready'
3. sol	sou-soal	'anchorage'
4. syop	sisop	'bathe'
5. tala	tela	'banana'
6. duil 'climb, as a tree'	dawil	'climb', ¹⁷
7. kal-uno	ulu	'leaf'
8. saya	saya	'flower'
9. galol	ailolo	'bush, forest'
10. snaŋa	sinana	'roast'
11. n- <u>uta</u>	ut	'bring'
12. lilo	pota-lil, nata-lil	'outside'
13. nipi	sani-nipa	'bee, wasp'
14. lafa	yafa-sa ¹⁸	'ten'
15. opiu 'day before yesterday'	p-oplu-ai	'day after tomorrow'
16. baŋsil	baŋseli	'flute', ¹⁹
17. ka-palo	pao	'half'
18. s-abe	g-abe	'in case, in the event that'
19. m-fis	fisa-n	'choose'
20. gala-n papo	ŋal-o	'chin'
21. ke- <u>nan</u>	ne-nena	'small'
22. pisi	pisi	'sick'
23. lukum ²⁰	lukam	'lansat (fruit)'
24. lek	le	'bad'
25. n-busu	busu	'lazy'

Misool	Buli	English
26. pit	bai-t (MD)	'moon'
27. falyan	failan	'mast'
28. fno	fano	'niece, nephew'
29. bilin	paliŋ	'stand up'
30. wili 'rope, rattan'	wala	'rope', ²¹
31. fatim ²²	fatan	'sniff, kiss'
32. g-mo-n	ohmo	'parents-in-law'
33. ka-	ki- (MD)	noun classifier? ²³
34. mo	moa	'to stream'
35. fayul	fayalo	'gather'
36. wanat-o	wanāt, wanāt-o (G)	'flesh'
37. ka-peo	pio	'fruit'
38. moro 'wind'	moda (G, MD)	'wind'
more-pat 'west wind'		
39. mit	met	'cloud'
40. im	em	'see'

Wallace (1869, Appendix) published a comparative vocabulary of 117 words in 33 languages of Indonesia. Two of these lists were collected for him on Misool. Wallace's Misool list 1 (no.49) most closely resembles that published by van Peski (1914), and can be taken as representing the same language. His Misool list 2 (no.50) differs from the former, and from van Peski's material in certain respects, as in having initial epenthetic *y* (*x₁apuy > yap 'fire'), next to epenthetic *l* in the other lists (Wallace list 1 lap 'fire', van Peski mlisin < *ma-qasin 'salty'). It seems clear, however, that Wallace's lists represent closely related languages, and the inclusion of both in the south Halmahera group thus appears justified.

Although to my knowledge no linguistic material has ever been collected from the islands of Ef Torobi and Kofiau (north of Misool), it is noteworthy that the former evidently bears in its name a word for 'island' that is known only among members of the south Halmahera group (Gane waf, Misool 1 yef, Misool 2 ef 'island'). If named by its own inhabitants, then, it seems likely that the population of at least the former island also speaks a south Halmahera language.

From the materials presented in Tables 5 and 6, and from the evidence that the phonological histories of Gane and Makian Dalam are substantially identical with that of Buli, it seems justified to conclude that the existence of a subgroup of AN languages comprising the languages of south Halmahera (with east Makian and Kayoa) and Misool is established.

It was suggested earlier that the ordering of the changes 1. -V > Ø and 2. *k > Ø as 1,2 in Buli, but 2,1 in Numfor is best interpreted as indicative of rule borrowing, with change 1 diffusing into Numfor from the west, and change 2 diffusing into Buli from the east. As noted above, one of the 13 sound changes shared by Numfor with Buli is not shared with Gane and Makian Dalam. This change is the loss of *k. More specifically, the available evidence suggests that medial *k disappeared in Gane and Makian Dalam (*iS(e)kan > ian 'fish'), but that initial and final *k have normally been retained (*kalibənəŋ > G kalibobo (Wallace), kalipép (Teljeur) 'butterfly', *kita > G kit (T) '1st pl. incl. actor', *kuCu > G, MD kut 'head louse', *kuDen > G kulan, MD kulaŋ 'earthen cooking-pot', etc.).²⁴ The south Halmahera languages of the southern dialect group thus offer striking empirical confirmation of an inference that was originally supported only by differences in the ordering of two phonological changes that both Buli and Numfor have undergone.

THE SARERA BAY LANGUAGE GROUP

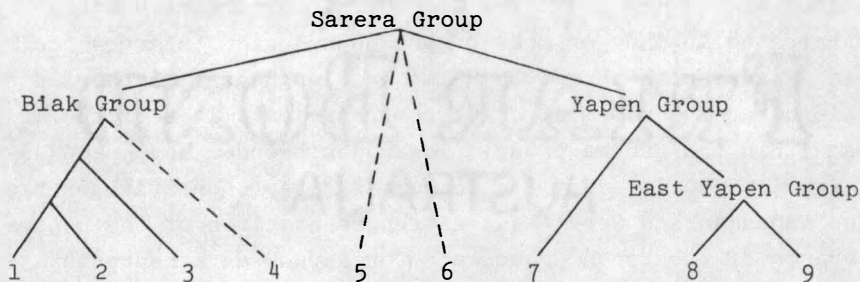
As noted earlier, Adriani and Kruijt (1914) stated that the languages of southern Halmahera subgroup with 'Numfor and its relatives'. They did not make clear what languages they would include among Numfor's 'relatives', though Esser (1938) specifically mentions Numfor, Windesi and Kowiai among the 'West New Guinea' languages.

In what is perhaps the most detailed classification yet made for the area, Anceaux (1961) considered materials for some 22 languages, of which 14 are spoken primarily on the island of Yapen, seven around the shores of Sarera Bay and on its small offshore islands, and one (Irarutu) on the narrow neck of land connecting Bintuni Gulf and Arguni Bay. Based on a combination of qualitative and quantitative evidence Anceaux reached the following major conclusions:

- 1) all of the languages considered except Irarutu appear to belong to a single subgroup of AN languages
- 2) this (unnamed) group divides into two fairly discrete first-order subgroups; one consisting of Biak, Ron, Dusner and probably the language of Meoswar island (for which data were insufficient to permit definite conclusions), the other including all the AN languages of the island of Yapen (exclusive of recent Biak settlements) and Wandamen-Windesi, spoken around Wandamen Bay, on the western side of Sarera Bay.

Anceaux called the first of these two co-ordinate groups the Biak group, but left the second group - like the total collection of subgrouped languages - unnamed. He noted, however, that Wabo and

Kurudu - spoken around the eastern end of the island of Yapen - diverge considerably from other Yapen languages, and placed them in a separate group which he called the 'East Yapen group'. It was added (pp.146-7) that "Waropen occupies a central position between Wandamen and the Biak group, and Mor is about half-way between Wandamen and Waropen." The wording of this passage apparently conflicts with that in the paragraph which follows (where it is stated that Wandamen-Windesi groups with the languages of central and western Yapen). Anceaux (p.c.), however, points out that 'Wandamen' in the above statement (and at several other places in his discussion), though not so indicated in the text, was intended to denote the lexicostatistically-defined subgroup which consists of Wandamen-Windesi and the AN languages of Yapen. Under this interpretation - and supplying the labels 'Sarera Group' for the entire collection of languages considered (exclusive of Irarutu), and 'Yapen Group' for Wandamen-Windesi and the AN languages of Yapen - the relations of these languages might be represented in a tree-diagram as follows:



1 = Biak, 2 = Ron, 3 = Dusner, 4 = Meoswar, 5 = Waropen, 6 = Mor, 7 = Wandamen=Windesi and all of the languages of Yapen and its satellites except the East Yapen group, 8 = Wabo, 9 = Kurudu

It is impossible to interpret Anceaux's statements regarding Waropen and Mor in such a way that they are compatible with a family tree model of linguistic differentiation, since a language B that is intermediate between A and C must in some ways be closer to A and in others closer to C, thereby leading to convergent branching. For convenience, in the following discussion each of these languages will be treated as a first-order branch of the Sarera Group.

As seen in section 1, Dyen (1965) combined Numfor and Biak in the Biakic Subfamily, and Wandamen and 'Yapen' (language unspecified) in the Wandamic Subfamily. He then united these two groups in the Geelvink Hesion. Based on what he calls a 'subadequate list' Dyen (p.45) also placed Kurudu in the Geelvink Hesion, but noted its low

percentages with all other members. Dyen's classification of most of the languages of Sarera Bay is thus in essential agreement with that of Anceaux. The chief point of difference between these classifications involves the position of Waropen. In Dyen's classification Waropen was originally regarded as highly discrete (in fact, as one of the 40 branches of the Austronesian Linkage). As observed earlier, however, in the appendix to his classification Dyen reports a restudy which suggests that Waropen and the languages of the Geelvink Hesion (there rechristened the Geelvink Subfamily) "can be associated directly with the Moluccan Linkage".

Comparative material consisting of about 250 words is available for all of the languages included in Anceaux's study. In addition, I have made use of the extensive Waropen wordlist of Held (1942), and of unpublished fieldnotes on Mor generously put at my disposal by D.C. Laycock.

If the subgrouping of Sarera Bay languages suggested by earlier writers is essentially correct, it should not be necessary to cite material from all of the languages in order to determine whether they have participated in many or most of the phonological changes that Numfor-Biak shares with the south Halmahera languages. Rather it should be sufficient to cite material only from representatives of the major subgroups within the Sarera Group. Toward this end I have made use of material from Anceaux's study for Dusner as a representative of the Biak Group, for Wandamen and Serui-Laut as representatives of the central-western branch of the Yapen Group, and for Kurudu as a representative of the East Yapen Group. Finally, I have examined the material for Irarutu to determine whether this language (and other languages in the Bintuni Gulf-Arguni Bay area which may subgroup with it) shows evidence of any of the phonological innovations shared by Numfor-Biak with the languages of the South Halmahera Group.

Once again, limitations of space prevent my stating all the available evidence here. The most noteworthy conclusions are nonetheless given below. All languages of the Sarera Group (including Waropen) show evidence of innovations 2, 4, 5, 6, 9 and apparently 12. As already observed, however, innovation 5 (*k, *q, ?, *H, S, x > Ø) is only partly attributable to changes in an immediate common ancestor, since the loss of non-medial *k is not shared with Gane and Makian Dalam. Evidence for the remaining changes is as follows (D = Dusner, War = Waropen, Wan = Wandamen, S-L = Serui-Laut, Kur = Kurudu, Ir = Irarutu):

2. *e₂ > South Halmahera, Misool, Sarera Bay, Irarutu o: *deŋeR > Mor -oran-i (met.), Ir b-nogr 'hear', *tebuS(u) > War kowu, Mor (k)-oha, Wan tobu, S-L tovu, Ir tof 'sugarcane', *telu > D tori, toru, War, Mor oro, Wan toru, S-L, Kur bo-toru, Ir tor(u), tur(ə) 'three'
4. *C,t,T/_i, *c, *-j-, *s > SH, Misool, SB, Irarutu s: *t-ina > Wan siña²⁵ 'mother', *tinaqi 'small intestine' > D sne, Mor sine,²⁶ Wan sane, Kur sine 'belly', *ŋajan > War nasan-o, Mor natan-(a) 'name', *pajey 'rice plant, rice in the field' > D pas, Wan fas, S-L fa 'rice', *si iDa > D si-i, Wan si, sia-t, Mor ti,²⁷ S-L sa, Kur i-si(a) '3rd pl. actor', *susu > D sus, War susi, Wan susu, S-L, Kur su, Ir sus 'breast'
6. *d,D, *z,Z, *l, *r > SH, Misool l, SB, Irarutu r:²⁸ *daSun > War ran-a (?), Mor ranu (met.), Wan rau, S-L re-rau, anda-rau(ŋ), Ir ro 'leaf', *DuSa > D su-ru, nu-ru, War wo-ru, Mor ru-ru, Wan mon-du, S-L, Kur bo-ru, Ir ru-(e) 'two', *Zalan > D rian,²⁹ War rara-do, Mor rarin-(a), Wan, Kur ran, S-L raŋ, Ir ran-dni ~ ra-d-n-i (met.) 'path, way, road', *lima > D rim-bi, War, Mor rim-o, Wan rim, S-L ri(ŋ), Kur bo-ve-rim '5', *lanjic > D ranet ~ rante (met.), Ir raŋəta 'sky'
9. *e₁, *a > SH, Misool, SB a, Ir Ø, e: *deŋeR > Mor -oran-i (met.) 'hear', Ir b-nogr 'hear', *inep > D enep, War ena-ko, Mor i-ena, Wan, S-L ena, Kur -ena, Ir g-in, ŋg-ine 'sleep', *ma-qitem > Wan meta(n), S-L nu-meta(ŋ), Kur mo-metr(a) 'black', *qatep > War aka, Mor r-aʔa 'thatch, roof', *tanem > War ana-ko, Wan tanam, S-L tana 'plant, bury'; *iS(e)kan > D in,³⁰ Mor r-ian, Wan d-ia, S-L d-ian, d-iaŋ, Kur d-ian, d-in 'fish', Ir d-ye 'flesh', *(s)a(R)man > War soman-o, Wan soma(n), S-L oman 'outrigger float'

Innovation (12) (Syncope) poses special problems that I can only touch on here. While syncope is widespread throughout the Sarera Bay languages, the vowel that is deleted in a given lexical item in most languages is retained in a few others, as with Buli mta, Nf mga, but Dusner mata ~ meta 'eye'. For the present it will be assumed that the interconsonantal vowel of forms such as Dusner mata ~ meta 'eye' is epenthetic. There is some support for this interpretation in the appearance of a for expected i in e.g. *tinaqi 'small intestine' > Wan sane, S-L ane 'belly', but other languages preserve the original quality of the vowel in question: Mor, Kur sine. In any case, since syncope is to some extent lexically specific, and since there is considerable agreement between the south Halmahera languages and the Sarera Bay languages in the forms affected, there can be no reasonable

doubt that a rule of Syncope was added to a language ancestral to all attested languages that show this change in certain lexical items.

The evidence for Syncope in Waropen is particularly striking. Waropen permits no initial consonant clusters, and normally retains the initial syllable of original disyllables and trisyllables. However, some initial syllables have been lost, and wherever the comparative evidence permits us to infer the cause of this otherwise unexplained development we find that the lost syllable of Waropen corresponds to the initial member of a consonant cluster which results from Syncope in other languages: *banua 'inhabited area' > Buli pnu, Numfor menu,³¹ War nu 'village', Buli ntu, POC *natu, War ku, ku-ku 'child'.

It seems incredible that the foregoing collection of phonological innovations shared by the south Halmahera languages with the languages of Sarera Bay could be a fortuitous product of convergent evolution. The development of *e alone (as penultimate o, final a) sets these languages off from all other members of the AN family, and could be taken by itself as subgrouping evidence to be reckoned with. But the evidence for a SHWNG subgroup is not confined to the five phonological innovations described in the preceding pages. There is in addition a substantial corpus of apparent lexical innovations, which will be listed in full in the revision of my conference paper. Moreover, various irregular phonological changes and some semantic innovations which appear to be shared exclusively by SHWNG languages provide powerful additional evidence in support of the South Halmahera-West New Guinea hypothesis. Some representative examples are as follows:

Shared irregularities in phonological development

1. *(ma)-Dalem > Buli m-laman, Gane m-loman, MD mu-laman, Misool malaman, Numfor ramen 'deep' (all with metathesis of the second and third consonants and irregular change *l > n)³²
2. *peñu > Buli, Gane fen, MD hen, War eni 'sea turtle' (all with e for expected o)
3. *si iDa > Buli, Gane, Numfor si, Dusner si-i, Waropen ki (for expected **sir, Mor ti, Wan si, sia-t, S-L sa, Kur i-si(a) (all with irregular loss of *ɔ) '3rd pl.'³³
4. *t-ina > Buli hñe, Wan siña (both with sporadic palatalisation) 'mother'³⁴

Semantic innovations

1. *qabaRa 'shoulder' > Buli pa, War awar-o 'carry on the shoulder'

Although the foregoing observations can hardly fail to carry considerable weight as subgrouping evidence, perhaps the single most interesting and powerful piece of evidence for a SHWNG subgroup is yet to be mentioned.

It was noted earlier that both Buli and Numfor have undergone the changes 1. $-v > \emptyset$ and 2. $*k > \emptyset$, but that the order in Buli was 1,2, and in Numfor 2,1. Wave theory as developed in European dialectology first came into conflict with the family tree model of linguistic differentiation because it recognised that diffusion is a possible source of innovation, and that isoglosses need not form discrete collections, but can - and frequently do - overlap. Anttila (1972) calls such areas of overlapping isoglosses 'transition areas', and he describes them (p.298) as follows:

Transition areas can be defined by the sharing of rules as well as by other items; that is, a dialect with rule 1 and another with rule 2 have a buffer zone with both rules 1 and 2. Such a geographical situation might even explain different rule order in a concrete fashion Suppose that rules 1 and 2 spread from opposite directions until they finally overlap. In the transition area closer to 1, rule 1 would have come earlier than rule 2, and vice versa. The dialects would now be defined by the following conglomerations: [1] — [1,2] — [2,1] — [2].

As can be seen, Buli and Numfor fit the classical model of a transition area. Moreover, in the extreme west of the SHWNG language area change 1 ($-v > \emptyset$) has occurred, but not change 2 ($*-k > \emptyset$). Finally, it is clear from the data presented above that a number of the eastern languages in the Sarera Group retain original final vowels in at least some words (e.g. '*sugarcane*', '*three*', '*mother*', '*breast*'). These languages have thus undergone change 2, but not change 1.

It is commonly believed that the diffusion of linguistic innovations is possible among dialects of the same language, but difficult or even impossible over distinct languages.³⁵ If this assumption is correct there would appear to be no escape from the conclusion that the SHWNG languages not only form a subgroup of the AN family, but that they also continue an ancient dialect chain.

Scope of the SHWNG subgroup

To determine the scope of the SHWNG subgroup I have made a preliminary examination of materials for the following languages or language areas that, in view of their geographical proximity to the area considered, seemed possible candidates for inclusion in the group:

1. Batjan, 2. Ambon-Ceram-Buru, 3. Kei, 4. Sula.

In no case have I found convincing evidence for including any of these languages in the SHWNG group. However, for Batjan I rely entirely on the statement and limited material supplied by Adriani and Kruijt (1914). This material suggests that the language/s of Batjan does/do not belong to the south Halmahera group, but the evidence is so limited that it leaves much room for reasonable doubt.

For Ambon-Ceram-Buru I have relied primarily on the classic work of Stresemann (1927), and here there can be no doubt that the phonological history is fundamentally different from that of the SHWNG languages, as seen in the following summary:

	SHWNG	Proto-Ambon
2. *e ₂	o	ø
4. *C, t, T/ _i	s	t
*c	s	s
*s	s	s
*-j-	s	l
6. *d, D	r	d
*z, Z	r	d
*l	r	l
*r	r	l
9. *e ₁	a	ø
*a	a	a
13. Syncope	+	-

Of the five phonological changes that define the SHWNG group, two are shared in part with Stresemann's Proto-Ambon: merger of *c with *s, and the double merger in Proto-Ambon corresponding to a single merger in Proto-SHWNG change 6 (*d, D with *z, Z, and *l with *r). Of these, only the merger of *d, D with *z, Z and the merger of *l with *r can be regarded as interesting, since the merger of *c with *s is shared widely with other AN languages, and therefore counts little as subgrouping evidence.³⁶

For Kei I have relied on Geurtjens (1921), and for the languages of the Sula islands I have taken the Soboyo vocabulary of Fortgens (1921). The full evidence for excluding these languages from the SHWNG group will be presented in the expanded version of this paper.

Culture-historical inferences

Space does not permit an in-depth exploration of the culture-historical inferences that can be justified by the subgrouping conclusions reached above, and by the reconstruction of the Proto-SHWNG vocabulary. However, it does appear that linguistic diversity is

greater among the languages of Sarera Bay than among the languages of southern Halmahera (though little is known of the intervening area). If the index of diversity in the two areas is a simple reflex of the normal process of differentiation it would seem to follow that speakers of SHWNG languages have been in Sarera Bay longer than in southern Halmahera. In fact, the relatively slight diversity among SH languages suggests that southern Halmahera has not been populated for any great length of time by Austronesian speakers. Since AN speakers would almost certainly have contacted this area in penetrating Melanesia, however, there is some reason to suspect that southern Halmahera has been subject to a good deal of language levelling. The present population of southern Halmahera thus may well represent a gradual westward expansion from the nearer end of an ancient dialect chain that was once confined to the northern Vogelkop Peninsula and Sarera Bay.

3. EVIDENCE FOR AN EASTERN MALAYO-POLYNESIAN SUBGROUP

The existence of a SHWNG subgroup of AN languages now seems to be reasonably well established. We can further conclude with some confidence that this subgroup was once a dialect chain in which the languages occupied approximately their present relative positions, though their absolute positions have probably changed through continuous expansion from Sarera Bay and the northern Vogelkop Peninsula toward the west.

Because of the time which it has taken to establish the SHWNG subgroup on fairly firm foundations, less time has been spent in testing the more inclusive Eastern Malayo-Polynesian hypothesis than was originally planned. Despite this limitation, and the fact that some of the evidence initially offered in support of the hypothesis has since been found to be invalid, the case for an Eastern Malayo-Polynesian subgroup comprising the SHWNG languages and the Oceanic languages has continued to grow in strength.

In contrast to the relatively strong phonological evidence for a SHWNG subgroup, there is little phonological evidence in support of an Eastern Malayo-Polynesian subgroup. There are, however, two pieces of evidence which might be cited as possible shared phonological innovations: the shift of penultimate *e to o, and the merger of *-j- with *s (and apparently *c) in both groups. The first of these changes would carry considerably greater weight if it were unconditioned in the SHWNG languages. Thus, PAN *e unconditionally yielded POC *o, but only penultimate *e yielded Proto-SHWNG *o. In view of the other evidence to be considered, it is perhaps simplest to assume that

penultimate *e first shifted to *o in a language ancestral to the SHWNG and Oceanic groups, and that last-syllable *e then followed this development in Proto-Oceanic, but merged with *a in Proto-SHWNG. Under this interpretation the change of PAN *e to POC *o would have occurred in two stages (first in the penult, then in the ultima), rather than as the single change that has generally been assumed. But we cannot entirely rule out the possibility that this change was independent in the two groups.

The merger of *-j- with *s is faced with similar problems, since in the available data *j in final position merged with *d, D, *z, Z, *l and *r (as in e.g. Proto-SHWNG *pusar 'navel'). Moreover, evidence has been presented elsewhere that despite its widespread merger with *s, *z, Z and probably *c in Oceanic languages, PAN *j apparently was retained as a distinct segment in Proto-Oceanic (cf. Blust to appear a). The existence of an Eastern Malayo-Polynesian subgroup thus cannot be based on the evidence of exclusively shared phonological innovations.

When we turn to a consideration of lexical evidence, and to some extent of grammatical evidence, however, we are confronted with a number of observations that *in toto* are difficult to explain unless we assume that the SHWNG and Oceanic languages have shared a period of common history apart from other AN languages. It will perhaps be useful if we first eliminate that evidence initially (viz. in Blust 1974) offered in support of the Eastern Malayo-Polynesian hypothesis which has since proved to be invalid. So far this consists of a single etymology, originally cited as follows:

category 1a

6. POC *Ropo, Buli opa, Numfor rob 'to fly'

It is now clear from the occurrence of Chamorro gupu 'fly, move through the air' that the Eastern Malayo-Polynesian forms in fact derive from a Proto-Malayo-Polynesian prototype *Rebek 'to fly'. Since we appeal only to exclusively shared innovations for subgrouping purposes, this item must be eliminated from the corpus of supporting evidence.³⁷

The evidence for an Eastern Malayo-Polynesian subgroup can be divided into three categories: 1) lexical innovations, 2) formal innovations (apart from regular sound change), in which category will be included sporadic phonological changes, loss of morpheme boundaries, etc., and 3) semantic innovations. The proposed lexical innovations will be presented first.³⁸

Possible Eastern Malayo-Polynesian lexical innovations

LI

1. PEMP *qanus-(i) 'to spit'
Serui-Laut k-unui (*a > u unexpl.), Wandamen k-anisu (met. of *kanusi) 'saliva', POC *qanus 'spittle', *qanus-i 'to spit'
2. PEMP *ayawan 'k.o. tree: Ficus sp.'
Buli yawan, Wogeo vaiawa, Gedaged aiau, Wuvulu aiwa 'a tree: the banyan'
3. PEMP *boi/bui 'smell, stench'
Buli pu-pui 'stench, stink', Raluana bu-bu 'smell offensively, have a bad or evil smell', Fijian boi 'have a smell',³⁹
4. PEMP *budan 'white'
Buli bu-bulan, POC *pulan 'white',⁴⁰
5. PEMP *butak 'close the eyes'
Buli bu-butak 'close the eyes', Roviana puta, Motu ma-huta 'sleep',⁴¹
6. PEMP *dadas-(i) 'peel off'
Buli lalas 'come off, of armbands, bandages, etc.', POC *dadasi 'scratch, peel, cut off'
7. PEMP *da(q)um 'shade, shady'
Numfor ka-daum 'shade', Raluana dau-dau 'shade, shadow', ka-dau-dau, ma-dau-dau 'shady'
8. PEMP *dege 'cavity, small recess'
Buli leg 'cave, grotto, hole', POC *ndeke 'hollow, concavity, small room or recess, pocket of seine net',⁴²
9. PEMP *ibu/ubi 'drinking vessel'
Buli ubi 'half coconut shell used for drinking cup', PPN *ipu 'container for liquid'
10. PEMP *iRiR 'to fan'
Numfor ir ~ rir 'fan oneself', POC *iRiR 'to fan',⁴³
11. PEMP *ka(dR)a 'cockatoo; parrot'
Ansus, Ambai kara, Serui-Laut kara-i, Blak kara-sibido⁴⁴ 'cockatoo', Roviana kara 'the general name for parrots'
12. PEMP *kiñit 'pinch'
Makian Dalam (de Clercq 1890) kinit, POC *kiñit 'pinch'

LI

13. PEMP *laman(a) 'deep sea'
Dusner ramen 'sea', POC *laman(a) 'sea, lake',⁴⁵
14. PEMP *ma- 'directional particle'
Buli lau 'sea-side' (opp. 'land-side'), ma-lau 'go toward the sea',
Hawaiian kai 'sea', ma-kai 'toward the sea', uka 'inland, upland',
ma-uka 'toward the interior'
15. PEMP *mada- 'prefix carrying a depreciatory force'
Buli mara- 'prefix of the accidental passive', Arosi mara- 'prefix
with depreciatory force',⁴⁶
16. PEMP *mada 'ripe, soft'
Buli mara 'ripe, cooked', Numfor mra 'light in colour; ripe, of
fruits which become lighter in colour when ripe', POC *mada
'fermented, soft, ripe, wet'
17. PEMP *maiduŋ 'evening, dusk'
Gane mailiŋ 'evening', Waropen mairon-o 'evening, sunset', Motu
mairu 'evening twilight',⁴⁷
18. PEMP *maluRu 'shade'
Buli ma-malu 'shadow (of things, not of people)', Raluana malur
'shade; shady', Fijian malu-malu 'shady', PPN *malu 'shade',⁴⁸
19. PEMP *matu 'dry coconut'
Numfor mgu 'hard, as nuts which harden when they are ripe', Mailu
matu 'hard, as a dry coconut', Lakona (Tryon 1976) matu, Dixon
Reef (Tryon 1976) na-mat 'dry coconut'
20. PEMP *mawa 'enclosed space'
Buli mawa 'roomy, spacious, as a house', Arosi mawa 'a cleft, space
between two rocks'
21. PEMP *mo(dl)aŋ 'real, genuine, true'
Buli molaŋ 'correct, real, genuine, true', Arosi mora 'original,
native, true, real, customary'
22. PEMP *momeq/momaw 'crumbs, sweepings, rubbish'
Buli mom 'rubbish, scraps', Numfor ai mom 'sawdust, wood chips',
Sa'a momo 'sweepings, rubbish', Arosi momo 'bit of fallen food',
Fijian momo 'break into small pieces', Samoan momo 'broken
remnants'

LI

23. PEMP *mumo 'whisper'
Buli fa-mu-mumo 'whisper to s.o.', Mota mum 'make indistinct hum'
24. PEMP *(n)a(dR)l 'to wait'
Numfor nari (retention of final vowel unexpl.), Motu nari 'to wait'
25. PEMP *natu 'child'
Buli ntu, Waropen ku, ku-ku, POC *natu 'child',⁴⁹
26. PEMP *(n)iwī 'nest'
Numfor niw 'nest (esp. a pig's lair)', Mor nibi ~ niwi, Waropen ni, Sa'a niui 'nest',⁵⁰
27. PEMP *ŋu(dR)um 'grunt, growl'
Numfor nurem 'grunting (of a pig), purring (of a cat)', Raluana ta-ŋuru 'grunt, as a pig', Sa'a ŋuru 'to growl or roar, of animals, to mumble or groan, of persons'
28. PEMP *patote(k,q)/patotaw 'outrigger boom'
Numfor fakok, POC *patoto 'outrigger boom'
29. PEMP *beke 'defecate'
Buli peke (retention of *k unexpl.), Raluana peke 'defecate'
30. PEMP *sala 'sharp-pointed object'
Numfor sar 'sharp', Nggela hala 'an echinus, sea-egg', PPN *tala 'sharp-pointed object'
31. PEMP *sakaRu 'reef'
Buli sa, POC *sakaRu 'reef',⁵¹
32. PEMP *sepat 'go past, go beyond'
Buli sefat 'go past, go beyond, go through', Arosi teha 'stray, wander', Bauro teha 'go past, beyond'
33. PEMP *sobu 'go down, descend; dive'
Buli sop 'go down, descend, dive', Numfor sop 'beneath',⁵² POC *sompū 'down'
34. PEMP *soRa/suRa 'to help'
Buli sua, POC *soRa 'to help'
35. PEMP *suda 'comb'
Waropen, 'Āre'Āre sura, Wuvulu cuka 'comb', Dobuan sura 'bunch of bananas',⁵³

LI

36. PEMP *supi 'peel, pare'
 Buli suf 'peel, pare', Numfor suf 'cut with a knife (of soft things), peel', Motu duhi-a 'pare, prepare (yams, etc.) for cooking', Gedaged supi 'pare, peel (cut off thin slices)', Arosi suhi 'peel yams', Fijian suvi 'cut in pieces, chiefly of yams and breadfruit, generally lengthwise'
37. PEMP *tabus 'taboo; sign of something interdicted'
 Numfor kabus 'taboo; tree branch or anything placed by the owner of a fruit tree, etc. near his property so that others will be afraid to approach it lest bad luck befall them', POC *tampu 'a ban, ritual restriction protected by supernatural sanction, marked by taboo sign'⁵⁴
38. PEMP *ta(dR)i 'steer a course (in navigating)'
 Numfor kar 'row (while facing one's destination)', Motu tari 'rudder, steer-oar; to steer a canoe'
39. PEMP *taŋa 'carrying container'
 Buli ta-taŋa 'any kind of container which one usually takes on a journey', POC *taŋa 'holder, bag'⁵⁵
40. PEMP *tatu 'true'
 Numfor kaku, Duke of York tatu-na 'true'
41. PEMP *tawan 'a tree: *Pometia pinnata*'
 Ambai, Wandamen tawa, Ansus tawan, Waropen kawan-o, Proto-Admiralty *taw(a), Fijian dawa 'a tree and its fruit: *Pometia pinnata*'
42. PEMP *tinan 'elder (of kinsmen)'
 Numfor sinan 'parents (of people and animals)', Proto-Admiralty *tinan 'elder (of siblings)'
43. PEMP *tobe(k,R) 'throw down'
 Buli topa 'throw down or away (as an anchor)', Arosi oho 'throw down'
44. PEMP *tobV 'fishnet float and the wood from which it is made'
 Buli tob 'wooden fishnet float', Selmat top 'tree with light wood used for fishnet floats', kaka-top 'fishnet float'

LI

45. PEMP *qutem/quten 'fishnet float'
 Buli utam 'fishnet float made from the mid-ribs of sago leaves',
 Wuvulu uʔo 'tree whose wood is used for fishnet floats, pillows,
 etc.', Raluana kutǎŋ, Bwaldoga utona 'fishnet float'
46. PEMP *qutub 'submerge to fill'
 Buli utup 'fill a bamboo with water in the river', Numfor uk (loss
 of last syllable unexpl.) 'pour water in a bamboo tube; carry water
 in a bamboo tube', POC *qutup 'flood, draw water, fill with liquid,
 soak'
47. PEMP *wa(q)ip 'scoop or container for water'
 Buli waif 'water scoop, bucket, pail', Arosi, Bauro wai 'coconut
 shell water bottle'
48. PEMP *watan 'bring, carry, take'
 Buli watan, Arosi wa 'bring, carry, take'

Possible Eastern Malayo-Polynesian formal innovations

FI

1. PEMP *kasu 'smoke'
 Numfor as, Wogeo kas, Manam ʔasu, Proto-Admiralty *kasu 'smoke',⁵⁶
2. PEMP *mutu 'broken off, severed'
 Numfor muk 'broken; break off, finish, of questions', PPN *mutu
 'cut off, ended'
3. PEMP *masuR/mosuR 'satiated'
 Buli mose, Windesi mos, Wuvulu maku, Aua maru, Fijian macu, Label
 masur, Raluana maur 'satiated',⁵⁷
4. PEMP *dui 'dugong'
 Waropen, Motu rui 'dugong',⁵⁸

Possible Eastern Malayo-Polynesian semantic innovations

SI

1. PEMP *kakat-(i) 'to peel (probably with the teeth)'
 Numfor ak, Windesi kati 'to peel', Raluana kat 'gnaw to bits; pull
 off the husk of a nut with the teeth', East Futunan kati 'remove
 the bark of something with the teeth',⁵⁹

SI

2. PEMP *lumut 'green'

Numfor rumek 'moss', man+dumek 'green', Dusner rumet 'green',
 Raluana limut 'the green colour or mossy growth on a canoe that
 has been standing in the water; green, blue, moss-green, colour
 of moss',⁶⁰

3. PEMP *miRmiR 'urinate'

Misool ta-mi (hence Buli fana-m 'urine', fana-mi 'urinate'), POC
 *mimiR 'spurt out, urinate',⁶¹

4. PEMP *pa(ka)-salaq 'punish'

Buli fa-sal, Numfor fa-sal, Samoan faqa-sala 'punish',⁶²

4. JUSTIFICATION OF THE LABEL 'EASTERN MALAYO-POLYNESIAN'

As with a small part of the evidence presented in my original research note (Blust 1974), some of the foregoing comparisons may well turn out to be chance resemblances, loans, or retentions. Of these three sources of potential error the problem of distinguishing innovations from retentions is undoubtedly the most difficult to control in the case of lexical comparison.

I have compared the 56 proposed PEMP innovations listed above with a number of other AN languages so as to minimise the probability that they are retentions, but time has not permitted systematic checking of all available sources, and further searching may well show that cognate forms or words with similar formal or semantic innovations occur outside the EMP subgroup. On the other hand it should also be pointed out that the magnitude of the task of justifying a South Halmahera-West New Guinea subgroup left me little time to search for additional Eastern Malayo-Polynesian lexical innovations. For this reason it is highly probable that many more forms shared exclusively by SHWNG languages with Oceanic languages will yet be found - especially when better dictionaries and wordlists become available for languages in the former group.

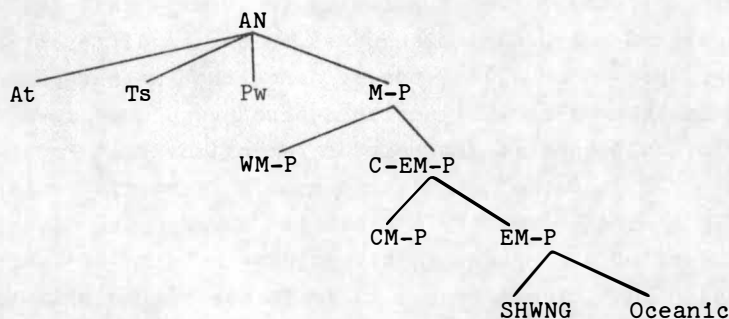
Finally, it should be mentioned that the problem of distinguishing innovations from retentions as I have discussed it in the preceding paragraph derives from uncertainty regarding the distribution of particular cognate lexical items rather than from uncertainty regarding the subgrouping of the languages compared. Stated somewhat differently, where the quantity of available evidence suggests that the subgrouping hypothesis adopted is likely to be justified in any event, the

determination whether a given word is an innovation or a retention must be made individually in each case. However, if it should happen that our subgrouping hypothesis were radically in error - say, that the SHWNG languages and the Oceanic languages belong to different first-order AN subgroups - then all of the evidence that we have adduced in favour of an Eastern Malayo-Polynesian subgroup would be *ipso facto* invalid, as it would consist entirely of retentions from Proto-Austronesian. The danger of committing such an error is particularly great if one proposes to justify a bipartite subgrouping division on the highest level of a language family solely in terms of lexical evidence, for under such circumstances it is impossible in principle to distinguish innovations from retentions. In this connection two general subgrouping principles are especially noteworthy: 1) in phonology - if we do not permit unconditioned phonemic splits - innovations can be distinguished from retentions, even on the assumption of two first-order subgroups, 2) a bipartite subgrouping can be justified solely on the basis of lexical evidence if the two groups involved in the bipartite division fall into a single first-order subgroup of the larger language family.

Assumptions are justified only to the extent that they serve to explain observations better than competing assumptions, and to my knowledge there are no qualitative observations which justify the assumption that the SHWNG languages and the Oceanic languages fall into different first-order AN subgroups. Moreover, a careful attempt to reconstruct the pronominal system of Proto-Austronesian (Blust 1977) indicates forcefully that all non-Formosan AN languages have undergone certain changes in the original pronominal system not known to be shared by any AN language of Formosa. In view of the fact that the non-Formosan AN languages in aggregate correspond closely to the collection of languages to which the term 'Malayo-Polynesian' was originally applied by Humboldt, it seems appropriate to revive this term as the designation of the non-Formosan group. As recognised by others before me, the Formosan languages themselves may fall into as many as three first-order AN subgroups (Atayalic, Tsouic, Paiwanic).

In my earlier research note I presented some evidence for a larger grouping of AN languages that includes Eastern Malayo-Polynesian together with various other languages of eastern Indonesia. In the interim time has permitted me to collect only a small quantity of additional evidence for such a grouping, but what evidence is available strongly suggests that further research will justify the recognition of a Central-Eastern Malayo-Polynesian subgroup. The subgrouping

relations that I recognise for the AN family on the highest levels are thus as follows:



At = Atayalic, Ts = Tsouic, Pw = Paiwanic, M-P = Malayo-Polynesian, WM-P = Western Malayo-Polynesian, C-EM-P = Central-Eastern Malayo-Polynesian, CM-P = Central Malayo-Polynesian, EM-P = Eastern Malayo-Polynesian

In summary then, the justification of the Eastern Malayo-Polynesian hypothesis depends

- 1) on evidence that all non-Formosan AN languages are members of a single first-order subgroup
- 2) on the existence of a body of lexical, formal and semantic innovations which - so far as has presently been determined - are shared exclusively by the SHWNG languages with the Oceanic languages
- 3) an appeal to principle 2 ('a bipartite subgrouping can be justified solely on the basis of lexical evidence if the two groups involved in the bipartite division fall into a single first-order subgroup of the larger language family'). Since the EMP subgroup is a bipartite genetic unit which does not involve different first-order AN subgroups, lexical innovations can be distinguished - at least probabilistically - from lexical retentions, and therefore exclusively shared lexical material can carry weight as subgrouping evidence.

N O T E S

1. Because it exceeded 130 typed pages and was still incomplete at the time of presentation, Dr Blust's conference paper will be revised and expanded for separate publication. The summary given here presents only an outline of the argument (originally sketched in Blust 1974), together with illustrative evidence.
2. Adriani and Kruijt cite no examples from Numfor. The van Hasselts (1947), however, list Numfor aibon 'fruit' (= ai 'tree' + bon 'fruit') and other forms which clearly establish that Numfor and the South Halmahera languages have the same genitive order in nominal compounds.
3. For an especially clear exposition of the procedure followed by Dyen cf. Grace 1966.
4. In his revised chart on p.58 Dyen calls the languages of the Geelvink group the 'Geelvink Heslon', but in his discussion on p.59 he refers to the same languages as the 'Geelvink Subfamily'. This apparent inconsistency evidently results from the Geelvink Heslon having first been revised to subfamily status (based on the recalculation of the internal percentages) then, following the recalculation with Waropen (averaged with all four Geelvink languages) returned once again to the status of a heslon.
5. Despite the lexicostatistical results obtained by Dyen both in his initial study and in his restudy of the languages of the Geelvink Heslon, it is generally agreed that Numfor and Biak are dialects of a single language. This oft-repeated claim is supported by Johsz Mansoben, a native speaker of Biak with whom I have been able to check the point.

6. There are no known Numfor reflexes of *c. The merger of *c with *s in Numfor nonetheless seems probable in accordance with the following reasoning: 1) *c was probably a voiceless palatal affricate, 2) in AN languages generally *c normally merged with some other segment or segments; in all known cases these segments include *s, 3) since Numfor lacks a voiceless palatal affricate it seems likely that *c has merged with some other PAN phoneme or phonemes, 4) given the general situation in AN languages the most likely candidate for merger with *c in Numfor is *s.

7. Hence *q- > Ø and *s- > Ø preceded and *k- > Ø (e.g. *kawil > B awil, Nf awir 'fishhook') followed y-epenthesis.

8. The only known Numfor reflex of *z contradicts the only known Numfor reflex of *Z: *zalateŋ > saraken 'stinging nettle, *Laportea sp.*'.

9. Although it normally reflects *b as b, Numfor occasionally reflects *b as p: *kalibenben > apop 'butterfly', and possibly *biRaS > pi 'roe, spawn', *tebel > kpor 'thick'.

10. Between contemporaneous descendants of a proto-language. King argues that ordering differences between successive stages of the same lineal tradition cannot be explained simply in this way.

11. It is noteworthy that Maan (1951:6) also calls the language of Gebe island a 'south Halmahera' language. This is Grace's 'Minyafuin', the language that Dyen (1965) found to be the closest relative in his sample to Buli.

12. Teljeur's material was obtained in the village of Gane Luar ('outer' Gane), the speech of which is said to be dialectally distinct from the speech of Gane Dalam ('inner' Gane). I wish to express my thanks to him for making available to me his unpublished fieldnotes, to Paul Alexander and Michael Young, then of the Australian National University, for collecting comparative wordlists for me from several parts of Halmahera in 1975, and to Carol Molony of Stanford University for making available a partially completed 200-item lexicostatistical test list for Makian Dalam.

13. Except that the change *a₁ > i (which is relatively rare in Buli) is not generally found among South Halmahera languages (e.g. *Zalan > B lalin, but Gane lolan, Makian Dalam lolan ~ lolan 'path, road').

14. *C,t,T/_i and *c are attested only in apparent loanwords (*timun > G timu (van der Crab), ka-tinum (Teljeur) 'cucumber', *timuR 'south-east monsoon' > G timur, MD timor 'south') and in *t-i(m)pu > G tibu (Teljeur), tibu-na (Maan 1951) 'grandparent', with apparently irregular *p > b. Pending the collection of additional data the development of these segments must, therefore, be regarded as indeterminate.
15. As reported by van Peski (1914). Two distinct lists appear in Wallace 1869, and will be discussed below.
16. As in Misool te-mtat, Maba ti-mtait < *ma-takut 'afraid'.
17. For the development of the sequence *-awi- cp. Misool uil < *kawil 'fishhook'.
18. 'One group of ten'; cp. Misool lafa-lu '20'.
19. Possibly a borrowing of Malay bansi 'flute'. If so, however, the liquid is distinctive.
20. Perhaps a printing error for lukam.
21. Presumably < PAN *waRej. What is of subgrouping interest here is the apparent metathesis of *R and *j.
22. Printing error for fatin?
23. As in Misool ka-pat, MD ki-pat < *batu 'stone'.
24. In a few words, however, initial *k has also disappeared: *kaSiw > G, MD ai 'wood', ai ai 'tree', *kaSu > G au, MD ao '2nd sg. actor'. Since intervocalic *k that came to be final as a result of the loss of a final vowel is retained (e.g. *aku > yak '1st sg. actor'), the change *-k- > ∅ evidently followed the loss of final vowels. Moreover, since the loss of final vowels, as we shall see, postdated the break-up of Proto-South Halmahera-West New Guinea, the loss of intervocalic *k in the South Halmahera languages and Numfor cannot be attributed to a single change in an immediate common ancestor, but must also have occurred independently in the two language groups. It appears, then, that *k > ∅ in intervocalic position spread first by diffusion, followed by the loss of non-medial *k, the latter change never reaching the western extremity of the original diffusion area.

25. Written *sinia*. I interpret the prevocalic sequence *-ni-* and the prepausal sequence *-in* in Anceaux's orthography as a palatal nasal. Thus *sinia* → /siña/ < *t-ina 'mother', Wan, S-L main → /mañ/ 'fat of an animal' < *meñak 'oil', etc.
26. Mor shows *C,t,T and apparently *c, *j, *s > s/ _i, but *C,t,T > Ø and *c, *j, *s > t elsewhere.
27. Also War ki (for expected **sir). This may have been borrowed from Mor before the change *t > k.
28. But d following a nasal.
29. Probably from earlier *rayan; cf. Pom rayan, Munggui layan 'path, way, road', with dissimilation of the sequence r ... r or r ... y.
30. With contraction of the vowel sequence by absorption of the second member, as in Numfor. Such contraction appears to be a drift phenomenon in the SHWNG languages, appearing also in Maba tos (expected **toas) 'ironwood' < *teRas 'core, heartwood of tree', and in Misool in < *iS(e)kan 'fish', o-fsa (Buli fasaw) < *pa(ka)-qa(s)awa 'marry', etc.
31. With unexplained *b > m. In the van Hasselts' orthography e after an initial consonant apparently is equivalent to zero.
32. An alternative etymology would associate this form with POC *laman(a) 'deep sea, sea beyond the reef'. If this interpretation is adopted the present comparison can be added to the evidence for an Eastern Malayo-Polynesian subgroup. However, on semantic grounds the association with PAN *Dalem 'deep' seems sounder.
33. Conceivably from PAN *si ia '3rd sg. actor'. If so, however, the change from singular to plural is found in all SHWNG languages. In Gane si may additionally be used as a polite singular, but this clearly is a secondary development from the plural meaning.
34. Possibly an independent development: cp. S-L mañ, Ansus miañ 'fat of an animal' < *meñak 'oil', where /ñ/ is found, but S-L, Ansus ina- 'mother', without a palatal. Also note the presumably independent development in Malay iña 'duenna', Javanese iña 'wet-nurse'.

35. This statement does not, of course, apply to surface linguistic features (as phoneme inventories, order of major constituents, use of prepositions vs. postpositions, etc.), since these are known to spread over distinct languages - even over languages which are not demonstrably related. But diffusion of this kind - which might be called 'typological diffusion' - is explainable by simple language contact. Since rule borrowing evidently must begin as lexical borrowing, the interpenetration of diffusing rules, on the other hand, is possible only where a substantial body of cognate vocabulary exists - that is, where sound correspondences are obvious to the naive native speaker. Without this precondition there is nothing which might enable the recipient community to identify a diffusing rule as a rule from the collection of lexical borrowings to which it is initially exposed.

36. To these we might add as a further feature separating the two groups the monophthongisation of diphthongs by assimilation and contraction in Proto-South Halmahera-West New Guinea (*-ay > e, *-aw > o, etc.), versus their retention, or monophthongisation by truncation in Proto-Ambon (*-ay > *-a(y), *-aw > *-a(w), *-uy > *-u(y), etc.). In addition, the merger of PAN *q,ʔ,H,S and x with zero is a further feature common to the two groups. Once more, however, the latter feature carries little weight as subgrouping evidence, since the merger of *ʔ,H,S and x with zero is almost universal outside Formosa and the Philippines, and the merger of *q with zero is sufficiently widespread to seriously compromise its value for subgrouping purposes.

37. In my original research note (Blust 1974), I pointed out that in addition to the evidence for a subgroup uniting the SHWNG and Oceanic languages, there is some lexical evidence for a larger subgroup including this and certain other languages of eastern Indonesia. One of the etymologies offered in support of this larger group was:

category 2c

3. Samoan maŋo, Buli maŋ (< PEMP *maŋ(o,aw)), Bimanese maŋo 'dry'

But the above forms, together with Manam raŋo, Raluana raŋ, maraŋ-(ana), Balangaw na-laŋo, Ata na-gaŋo can be attributed to PMP *(ma)-Raŋaw 'dry', and therefore must also be discounted as subgrouping evidence.

38. Proto-Eastern Malayo-Polynesian forms probably require the reconstruction of an oral grade/nasal grade distinction, but this is not attempted here. Partly through attrition as a result of my own

continuing research and partly through criticism of this section generously offered by Isidore Dyen, four lexical innovations, four formal innovations and one semantic innovation that were proposed in the conference paper have been dropped. It is hoped that a fuller response to Professor Dyen's criticisms and a more concerted attack on the oral grade/nasal grade problem will be possible in the envisaged expansion of my paper.

39. Dempwolff (1938) attempted to derive Fijian *boi* from PAN *baSuq '*stench, odour*', but two distinct roots evidently are needed to account for the facts (thus Buli *pau* '*scent, fragrance, aroma*' < *baSuq). The correspondence of *o* in Oceanic languages to *u* in non-Oceanic cognates - though unexplained - is found in a number of other forms.

40. Cf. also Iban *burak*, Roti *fulak* '*white*', with a different final consonant. As noted in an earlier work (Blust 1970), Nggela *pura* '*white*' suggests that the Proto-Oceanic reconstruction should be *pudan.

41. Evidently distinct from Dempwolff's *buta '*blind*' (cf. also Pailwan *ma-vutsa* '*have bad eyesight*').

42. As observed in my original research note, the PEMP *g/k distinction is tentative. Buli *gokgok* '*a bird: the crow*' < PAN *gekgek '*throaty cry*' may provide evidence of a second instance of PEMP *g, though its onomatopoeic character renders it suspect. It is possible that Buli *g* in non-onomatopoeic words derives from *ŋk.

43. This reconstruction appears as *iriR in Grace (1969). Tongan *ṽ* '*to fan*' supports the revision.

44. In the special language of dirges.

45. This reconstruction appears as *laman in Grace 1969. Since original final consonants and any preceding vowel have been lost in all languages of the Admiralty Islands, however, the occurrence of Loniu, Pak *laman*, Papitalai *lo-loman* '*deep blue sea, sea beyond the reef*' is most simply explained on the assumption that the last nasal in this word was originally followed by a vowel. Nonetheless, other languages in the same group support *laman: Penchal (Rambutyo) *lam* '*deep blue sea beyond the reef*', Wuvulu-Aua *lama* '*sea (in general)*'. Mussau *lamana* '*sea near the shore*', Roviana *lamana* '*the ocean; deep, of water*' can be reconciled with either shape of the reconstruction.

46. The basis of this cognate association is the observation that at least one class of agentless passive constructions carries negative emotional overtones for the experiencer in many languages (hence the only semi-facetious reference to this type of construction by some recent writers as the 'paranoid passive'). Examples are English 'get' passives and e.g. Bahasa Indonesia ke-hujan-an 'caught in the rain', ke-tidur-an 'oversleep', ke-besar-an 'too large (of shoes, clothes, etc.)', where the meaning of the root ('rain', 'sleep', 'large') is consistently coloured in ways that the experiencer is likely to find at least mildly unpleasant.

47. Buli maliŋliŋ 'afternoon', Makian Dalam maliliŋ 'evening' suggest that Gane has irregularly lost the first *l in this word. Since Motu has also lost intervocalic *l in some forms (*bulan > hua 'moon', *bulu > hui 'body hair, feathers'), it is possible that the entire set of words derives from *maliduŋ. If so, a connection with Philippine forms such as Kayapa Kallahan qalliduŋ 'shadow' is conceivable (the EMP forms thus deriving from *ma-(q)aliduŋ). In this event the EMP innovation would be semantic rather than lexical. However, Waropen mairon-o supports the Gane reflex in pointing to *maiduŋ.

48. The Oceanic forms could derive from *maluR, but the last vowel of the Buli word is incompatible with such a prototype (cf. *busuR > Buli pusi 'bow'). All items compared here, on the other hand, can regularly reflect *maluRu. As with most other trisyllabic reconstructions that begin with the phoneme sequence *ma-, this word probably contains the stative/attributive prefix *ma-.

49. Milke (1968) associated reflexes of POC *natu with non-Oceanic forms such as Ilocano nato 'ovary of a bird', Bolaang-Mongondow natu 'egg', Timorese ika in natu-n 'roe, spawn', etc., but claimed that reflexes of this item with the meaning 'child' are unknown in non-Oceanic AN languages. This comparison can be taken as evidence that Milke's claim is incorrect, and that *natu almost certainly replaced the reflex of PAN *aNak 'child' in Proto-Eastern Malayo-Polynesian. Isidore Dyen (p.c.) has suggested Malay menantu 'son (or daughter)-in-law' as a possible reflex of *natu, but under this interpretation both the morphology and the meaning of the Malay form present problems. In view of Minangkabau binantu, Merina vinantu 'son- or daughter-in-law'

this item is probably better regarded as a derivative of *bantu 'help, assistance' (*b-in-antu), as originally proposed by Dempwolff.

50. Sa'a niui may reflect POC *ñikut with the addition of an unexplained element -i after the loss of the final consonant.

51. Its phonetic similarity notwithstanding, Sangir saghe 'reef' apparently is not connected with these words.

52. The development of PAN prenasalised medial obstruents in Buli is still unclear. It is possible that *b, *mb and *mp all yielded Buli p. If so the present reconstruction contains an unindicated ambiguity for *(bp). For the Numfor reflex cf. note 9.

53. Despite their similarity to reflexes of PMP *(s)aRu 'comb', these forms appear to reflect a distinct root. Other phonetically similar, but distinct words are widespread in the Philippines, as with Western Bukidnon Manobo, etc. suwat, Ivatan sorod, Itbayaten surud 'comb'.

54. To my knowledge no pre-POC etymon has ever been proposed for this classic Oceanic term, which almost certainly replaced the reflex of PAN *palisi 'taboo' in Proto-Eastern Malayo-Polynesian. Reflexes in other SHWNG languages and in Oceanic languages which retain *-s are still unknown, but this gives no reason to doubt the validity of the etymology, as both the form and meaning are free from comparative problems.

55. Dempwolff connected reflexes of this word with Malay tañan 'hand' and related forms in other non-Oceanic AN languages, noting the agreement of the thematic consonant in Fijian taña 'bag, pocket, sack', taña-na 'put into a bag; fish with hand-nets' with the assumed historical final. Because Buli preserves original final *n, however, Buli ta-taña cannot point to earlier *tañan. If adopted this proposed cognate set thus suggests that the -n- of Fijian taña-na is a secondary development (cp. also taña-va 'catch fish in a hand-net').

56. Next to Dempwolff's *qa(s)u 'smoke'. As in Buli, initial a in Numfor can only derive from the earlier sequence *ka- (initial *a, or *a preceded by a laryngeal yielded ya-).

57. To account for the agreement of Tagalog *busóg*, Toba Batak *bosur*, Ngaju Dayak *besoh* 'satiated' Dempwolff reconstructed **be(s)uR* 'satiated'. In addition he posited a doublet **ba(s)uR* solely to explain the similarity of Fijian *macu* 'satiated', Sa'a masu 'vomit (of an infant cloyed on breast-milk)' to the first reconstruction. The appearance of m-initial forms in other, distantly related Oceanic languages (as Wuvulu maku, Aua maru 'satiated') clearly justifies the reconstruction of POC **masuR*, though direct evidence for a PAN form **ba(s)uR* is still to be sought. Buli *mose*, Windesi *mos*, on the other hand, appear to reflect **be(s)uR* with sporadic nasal substitution, though Oceanic forms of the type *mosur* are yet to be found. The agreement of SHWNG languages and Oceanic languages in showing sporadic nasal substitution in this form is thus parallel rather than direct, though a development PMP **be(s)uR* > PEMP **mosuR* (with sporadic nasal substitution) > POC **masuR* (with sporadic change of the penultimate vowel) is not inconceivable.

58. From Dempwolff's **Duyun*. Waropen normally retains final nasals (PEMP **maidun* > *mairon-o* 'evening, sunset', PEMP **kuden* > *uran-o* 'earthen cooking-pot'), and Motu retains the vowel after **y* in **layaR* > *lara* 'sail'. The development **Duyun* > *rui* in both languages thus appears to involve shared sporadic changes that have not been reported from any non-EMP language.

59. Non-EMP cognates such as Atta *man-gaga:q* 'bite', Kankanay *gatgat* 'crunch, munch, champ, chew', Karo Batak *gatgat* 'chop up fine, as earth with a hoe; mince' may refer to crushing or cutting with the teeth, but no known non-EMP cognate refers to peeling with the teeth. For the connection of reduplicated monosyllables in some languages with the unreduplicated root followed by **-i* in others cf. Blust to appear b.

60. From Dempwolff's **lumut* 'moss, algae'. The meaning 'moss' clearly persisted in Proto-Eastern Malayo-Polynesian as well, but evidently gave rise to a derivative colour term that is not known elsewhere in the AN family. The association 'moss; green' is also reported for *Kakiduge:n* Ilongot, but in connection with a different root (*guput*).

61. From Dempwolff's **mirmir* 'moisten, spray'.

62. In the limited time available between the writing of my conference paper and the preparation of this summary it has not been possible for me to check the meaning of this morphologically complex form in more

than a small number of languages. Moreover, the original meaning of *pa(ka)-salaq is still unclear. Nonetheless, the meaning '*punish*' has not yet been found in association with a reflex of *pa(ka)-salaq in any non-EMP language.

BIBLIOGRAPHY

- ADRIANI, N. and Alb. C. KRUIJT
 1914 *De Bare'e-sprekende Toradja's van midden-Celebes*, vol.3.
 VKNA 46. Batavia: Landsdrukkerij.
- ALEXANDER, Paul and Michael YOUNG
 n.d. Comparative Word Lists of Halmahera Languages. Typescript.
- ANCEAUX, J.C.
 1961 *The Linguistic Situation in the Islands of Yapen, Kurudu, Nau and Miosnum, New Guinea*. VKI 35.
- ANTTILA, Raimo
 1972 *An Introduction to Historical and Comparative Linguistics*.
 New York: Macmillan.
- BLUST, Robert A.
 1970 'Proto-Austronesian Addenda'. *OL* 9/2:104-62.
 1974 'Eastern Austronesian: a Note'. *WPLUH* 6/4:101-7.
 1977 'The Proto-Austronesian Pronouns and Austronesian Sub-
 grouping: A Preliminary Report'. *WPLUH* 9/2:1-15.
- to appear a 'The Proto-Oceanic Palatals'. To appear as Memoir 43 of
 The Polynesian Society, Auckland.
- to appear b 'A Rediscovered Austronesian Comparative Paradigm'. To
 appear in *OL*.

CHAFE, Wallace L.

1968 'The Ordering of Phonological Rules'. *IJAL* 34/2:115-36.

CLERCQ, F.S.A. de

1890 *Bijdragen tot de kennis der Residentie Ternate*. Leiden: Brill.

CRAB, P. van der

1862 *De Moluksche eilanden*. Batavia: Lange.

DEMPWOLFF, Otto

1920 *Die Lautentsprechungen der indonesischen Lippenlaute in einigen anderen austronesischen Südseesprachen*. ZES Supplement 2.

1924-25 'Die L-, R- und D-Laute in austronesischen Sprachen'. ZES 15:19-50, 116-38, 223-38, 273-319.

1934-38 *Vergleichende Lautlehre des austronesischen Wortschatzes*. ZES Supplements 15 (1934), 17 (1937), 19 (1938).

DYEN, Isidore

1965 *A Lexicostatistical Classification of the Austronesian Languages*. *IJAL* Memoir 19.

ESSER, S.J.

1938 'Talen'. In: *Atlas van tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

FORTGENS, J.

1921 *Bijdrage tot de kennis van het Sobojo (eiland Taliabo, Soela-groep)*. The Hague: Martinus Nijhoff.

GEURTJENS, H.

1921 *Woordenlijst der Keieesche taal*. VBG 63/3.

GRACE, George W.

1955 'Subgrouping of Malayo-Polynesian: a Report of Tentative Findings'. *AmA* 57:337-9.

1961 'Austronesian Linguistics and Culture History'. *AmA* 63: 359-68.

- 1966 'Austronesian Lexicostatistical Classification: a Review Article'. *OL* 5/1:13-31.
- 1968 'Classification of the Languages of the Pacific'. In: A.P. Vayda, ed. *Peoples and Cultures of the Pacific: an Anthropological Reader*, 63-79.
- 1969 'A Proto-Oceanic Finder List'. *WPLUH* 1/2:39-84.
- 1971 'Notes on the Phonological History of the Austronesian Languages of the Sarmi Coast'. *OL* 10/1:11-37.
- HASSELT, J.L. and F.J.F. van HASSELT
- 1947 *Noemfoorsch Woordenboek*. Amsterdam: J.H. de Bussy.
- HELD, G.J.
- 1942 *Woordenlijst van het Waropensch*. Bandoeng: A.C. Nix.
- KING, Robert D.
- 1969 *Historical Linguistics and Generative Grammar*. Englewood Cliffs, N.J.: Prentice-Hall.
- LAYCOCK, D.C.
- 1965-66 Unpublished fieldnotes on Mor. Manuscript.
- MAAN, G.
- 1940 *Boelisch-Nederlandsche woordenlijst ...* VBG 74/3. Bandoeng: A.C. Nix.
- 1951 *Proeve van een Bulische spraakkunst*. VKI 10.
- MASINAMBOW, E.K.M.
- 1972 'Halmahera'. In: Frank M. LeBar, ed. *Ethnic Groups of Insular Southeast Asia*, 119-22. New Haven: Human Relations Area Files Press.
- MILKE, Wilhelm
- 1958 'Zur inneren Gliederung und geschichtlichen Stellung der ozeanisch-austronesischen Sprachen'. *ZEthn* 83:58-62.
- 1968 'Proto-Oceanic Addenda'. *OL* 7/2:147-71.

MILLS, Roger

n.d. Index to Stresemann, 1927: *Die Lauterscheinungen in den ambonischen Sprachen*. Typescript.

MOLONY, Carol H.

n.d. Comparative Wordlists of Halmahera Languages. Typescript.

PESKI, F. van

1914 *Beschrijving eener Exploratie van het eiland Misool, met annex woordenlijst der taal*. Deventer: Velders.

STRESEMANN, Erwin

1927 *Die Lauterscheinungen in den ambonischen Sprachen*. ZES Supplement 10.

SUPARNO

1975 *Kamus bahasa Biak-Indonesia*. Report of the Results of Lexicographic Research, Project for the Development of the National and Areal Languages and Literatures of Indonesia (ILDEP). Jayapura: Universitas Cenderawasih.

TELJEUR, Dik

n.d. *Gane Luar-Nederlands woordenlijst*. Typescript.

TRYON, D.T.

1976 *New Hebrides Languages: an Internal Classification*. PL, C-50.

WALLACE, Alfred R.

1869 *The Malay Archipelago*. 2 vols. London: Macmillan.