THE LANGUAGES OF THE NEW HEBRIDES:
INTERNAL AND EXTERNAL RELATIONSHIPS

D.T. TRYON

1.0. INTRODUCTION

The Anglo-French Condominium of the New Hebrides consists of an incomplete double chain of more than 80 islands, of which 11 or 12 may be considered major islands. The New Hebrides archipelago stretches from north-west to south-east for approximately 550 miles, and lies between 12 and 20 degrees S.Lat. and 165 and 170 degrees E.Long. For administrative purposes the Banks and Torres Islands are considered part of the New Hebrides.

The languages of the New Hebrides are still very poorly known compared with the languages of other Austronesian areas, especially Polynesia and Micronesia. Until the last decade, published information on the languages of the group went back to the early studies of such scholars as von der Gabelentz (1861 and 1873), Codrington (1885), Kern (1906) and Ray (1926). More recent information was provided by Capell (1954 and 1962).

During the last decade there has been a considerable upsurge in interest in the languages of the New Hebrides. Studies of individual languages to appear in recent years include those by Walsh (1962 and 1966), Kasarhérou (1962), Hewitt (1966), Parker (1968 and 1970), Schütz (1969a and 1969b), Paton (1971 and 1973), Guy (1974), Charpentier (1974) and Lynch (1974). Depth studies are currently in progress on Paama (T. Crowley), Epi (D.T. Tryon), South Malekula (J.M. Charpentier), while Lynch is continuing his study of the languages of the three southern islands.

Since Capell's (1954 and 1962), there have been few comparative studies covering any large area of the New Hebrides, apart from those of the present writer (Tryon 1972, 1973 and 1976). In the latest

publication (Tryon 1976) an internal classification of all of the known
New Hebrides languages was presented, the main results of which are
summarised below.   Lynch (1975 and 1976, and this volume) has provided
considerable subgrouping evidence for the languages of the three south-
ern islands.

As far as the external relationships of the New Hebrides languages
are concerned, previous attempts have been made by Grace (1955), Dyen
(1965) and Pawley (1972).   While the studies of these three scholars
will be discussed below, it should be stated here that the languages
selected by them for subgrouping on an Oceania-wide or Austronesia-wide
basis were chosen mainly because materials in the languages used were
available, not because the languages were truly representative of New
Hebrides languages.   This shortcoming has been overcome in the present
paper, since after lengthy fieldwork the writer was able to complete an
internal subgrouping of all New Hebrides languages for the first time
(Tryon 1976), and was thus in a position to select representative lan-
guages when the external relationships of the languages of these islands
were considered.

It must be stressed, however, that while the internal subgroupings
established have been based on a study of sound correspondences and
lexicostatistical methods, requiring corroboration from the comparative
morphological study at present in train, the external relationships
suggested in this paper are based on quantitative evidence alone and
must be seen as a first approach to the subject only.   As the languages
both inside and outside the New Hebrides become better known, more
definitive progress in situating the New Hebridean languages within the
Oceanic and/or Austronesian framework can be anticipated.

2.0.   INTERNAL RELATIONSHIPS

2.1.   *QUANTITATIVE EVIDENCE*

Between 1969 and 1974 some 350 wordlists each of 309 items were
collected, mainly by the writer, representing what is believed to be
all of the extant languages of the New Hebrides.   For the computation
of cognate percentages the number of the lists was reduced from 350 to
178, thus eliminating lists which differed only minimally from each
other.   The number of items on the lists was reduced from 309 to 258
on the basis of difficulty of accurate elicitation in the Island
Melanesia region.   (For details see Tryon 1976:96ff.)

All of the lists used in the computation of cognate densities in-
corporated the Swadesh 100 and 200 wordlists as modified by Samarin
(1967:220-3), together with additional items considered suitable for
this geographical area.

Cognate identification was made on the basis of regular sound corres-
pondences with the Proto-Oceanic phoneme inventory as established by
Grace (1969), together with the addition of *ñ, as shown necessary by
Blust (forthcoming). The method of cognate recognition adopted may be
described as conservative, since in addition to reference to regular
sound correspondences, doubtful cognates were scored as non-cognate.

For purposes of the internal lexicostatistical classification the
following percentages were taken as diagnostic:

    Approximately 81%-100% Dialects of same language
    Approximately 50%-80%   Different language, same Subgroup
    Approximately 30%-49%   Different Subgroup, same Group
    Approximately 20%-29%   Different Group, same 'Family'

The rationale behind the subgrouping percentage cut-offs warrants brief
comment, as follows:

(a) 81% was chosen as the demarcation line between language and dialect
on the basis of observed mutual intelligibility; any lower figure proved
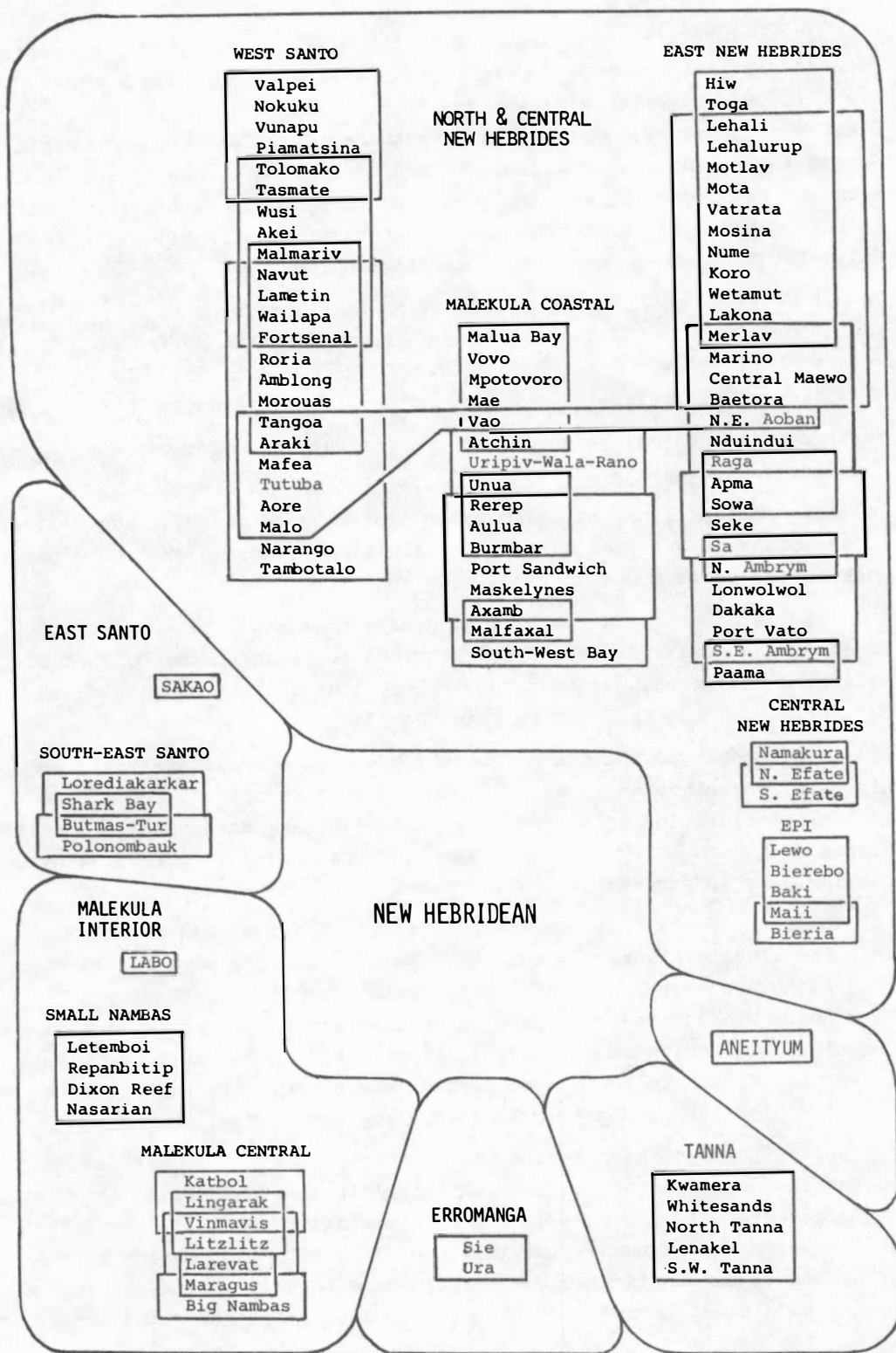unsatisfactory in a New Hebrides-wide context.

(b) 50% was chosen to delimit languages belonging to the same subgroup,
since it was observed, in the New Hebrides context, that some sort of
natural division occurs about this point, languages tending to score
several points higher or lower than the 50%.

(c) It was observed that languages which scored approximately 30% or
higher with each other appeared to demonstrate structural and morpho-
logical similarities with each other, whereas languages whose percentile
scores fell much below this mark were accompanied by a corresponding
falling away in structural similarities.

(d) The term 'Family' in the internal classification does not imply
that individual languages within the New Hebrides are more closely
related to each other than to any language grouping outside the New
Hebrides. Within the first part of this paper it simply places a geo-
graphical limitation on the languages considered. The matter of 'Family'
is taken up again in the second part of the paper, when the external
relationships of New Hebrides languages are considered.

(e) For purposes of the classification adopted, it was thought useful
to use the principle of 'dialect chaining' in the New Hebrides context.
A dialect chain is a series of speech communities such that the speech
of Community A is mutually intelligible with that of Community B, that
of B with C, but not A with C, setting up an intelligibility chain,
termed a 'dialect chain'. In the New Hebrides this system has been

CHART I: NEW HEBRIDES LANGUAGE CLASSIFICATION

WEST SANTO

Valpei
Nokuku
Vunapu
Piamatsina
Tolomako
Tasmate
Wusi
Akei
Malmariv
Navut
Lametin
Wailapa
Fortsenal
Roria
Amblong
Morouas
Tangoa
Araki
Mafea
Tutuba
Aore
Malo
Narango
Tambotalo

NORTH & CENTRAL NEW HEBRIDES

MALEKULA COASTAL

Malua Bay
Vovo
Mpotovoro
Mae
Vao
Atchin
Uripiv-Wala-Rano
Unua
Rerep
Aulua
Burmbar
Port Sandwich
Maskelynes
Axamb
Malfaxal
South-West Bay

EAST NEW HEBRIDES

Hiw
Toga
Lehali
Lehalurup
Motlav
Mota
Vatrata
Mosina
Nume
Koro
Wetamut
Lakona
Merlav
Marino
Central Maewo
Baetora
N.E. Aoban
Nduindui
Raga
Apma
Sowa
Seke
Sa
N. Ambrym
Lonwolwol
Dakaka
Port Vato
S.E. Ambrym
Paama

EAST SANTO

SAKAO

SOUTH-EAST SANTO

Lorediakarkar
Shark Bay
Butmas-Tur
Polonombauk

MALEKULA INTERIOR

LABO

SMALL NAMBAS

Letemboi
Repanbitip
Dixon Reef
Nasarian

MALEKULA CENTRAL

Katbol
Lingarak
Vinmavis
Litzlitz
Larevat
Maragus
Big Nambas

NEW HEBRIDEAN

CENTRAL NEW HEBRIDES

Namakura
N. Efate
S. Efate

EPI

Lewo
Bierebo
Baki
Maii
Bieria

ANEITYUM

ERROMANGA

Sie
Ura

TANNA

Kwamera
Whitesands
North Tanna
Lenakel
S.W. Tanna

carried a step further, as there are chains of languages such that
language A shares a relatively high percentage of common vocabulary
with language B, B with C, but not C with A, where the percentage is
considerably reduced. In the diagrammatic representation of the New
Hebrides 'Family' presented in Chart I, the chaining principle will be
seen to have been applied throughout.

In the East New Hebrides Subgroup, for example, languages A, B, C,
D, E, F, G, H, I, J, K, L, M all form a single subgroup on the basis
that they all share more than 50% cognates, based on the test list used
for the computation. However, languages C, D, E, F, G, H, I, J, K, L,
M, N, O, P also form a subgroup, as do M, N, O, P, Q, R, S. It will
be seen, then, that a system of overlaps is operating in the New
Hebrides languages. This overlap or 'subgroup chaining', as it might
be termed, goes some way towards resolving the problem of sharp cut-
offs between groups, and appears to answer to a sociological reality.

In the diagrammatic representation of the New Hebrides internal
relationships presented in Chart I, the chaining phenomenon observed
within the larger subgroups has been indicated by overlapping rectangles.
More than 50% cognates are shared by languages within any rectangle.

A perusal of the chart will show that Vao, in the Malekula Coastal
subgroup, acts as a link-pin between the West Santo, Malekula and East
New Hebrides subgroups. Such a feature is not unexpected, given the
chaining observed at the dialect, language and subgroup levels.

The lexicostatistical classification of New Hebrides languages, set
out in the traditional form, follows. The overlaps in subgrouping
illustrated in Chart I are not repeated in the classification given
below. For tables of percentages of shared cognates, see Tryon 1976:
95-162, where they are set out in full.

The New Hebrides 'Family'

1. East Santo Group (5 languages)

Sakao Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Sakao	E.Santo	1000

South-East Santo Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Lorediakarkar	E.Santo	50
Shark Bay	E.Santo	150
Butmas-Tur	E.Santo	350
Polonombauk	E.Santo	50

2. Malekula Interior Group (12 languages)

Small Nambas Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Letemboi	S.Malekula	182
Repanbitip	S.E.Malekula	60
Dixon Reef	W.Malekula	50
Nasarian	W.Malekula	20

Malekula Central Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Katbol	E.Malekula	250
Lingarak	E.Malekula	120
Vinmavis	W.Malekula	140
Litzlitz	E.Malekula	200
Larevat	W.Malekula	100
Maragus	N.E.Malekula	10
Big Nambas	N.W.Malekula	1200

Labo Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Labo	S.W.Malekula	350

3. Erromanga Group (2 languages)

Erromanga Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Sie	Erromanga	600
Ura	Erromanga	10

4. Tanna Group (5 languages)

Tanna Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Kwamera	Tanna	1100
Whitesands	Tanna	2500
North Tanna	Tanna	2000
Lenakel	Tanna	3000
South-West Tanna	Tanna	1600

5. Aneityum Group (1 language)

Aneityum Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Aneityum	Aneityum	320

6. North & Central New Hebrides Group (77 languages)

East New Hebrides Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Hiw	Torres Is.	50
Toga	Torres Is.	150

LANGUAGE	LOCATION	SPEAKERS
Lehali	Ureparapara Is.	100
Lehalurup	Ureparapara Is.	60
Motlav	Mota Lava Is.	850
Mota	Mota Is.	270
Vatrata	Vanua Lava Is.	100
Mosina	Vanua Lava Is.	400
Nume	Gaua Is.	120
Koro	Gaua Is.	70
Wetamut	Gaua Is.	70
Lakona	Gaua Is.	80
Merlav	Merig/Merelava Is.	850
Marino	N.Maewo	90
Central Maewo	C.Maewo	350
Baetora	S.Maewo	500
N.E.Aoban	N.E.Aoba	3000
Nduindui	W.Aoba	3000
Raga	N.Pentecost	2300
Apma	C.Pentecost	3000
Sowa	C.Pentecost	20
Seke	C.Pentecost	200
Sa	S.Pentecost	1200
North Ambrym	N.Ambrym	1900
Lonwolwol	S.W.Ambrym	400
Dakaka	S.W.Ambrym	400
Port Vato	S.W.Ambrym	500
South-East Ambrym	S.E.Ambrym	1000
Paama	Paama Is.	2000

West Santo Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Valpei	N.W.Santo	200
Nokuku	N.W.Santo	160
Piamatsina	N.W.Santo	150
Vunapu	N.W.Santo	100
Tolomako	N.Santo	350
Tasmate	W.Santo	100
Wusi	W.Santo	170
Akei	S.W.Santo	650
Malmariv	C.Santo	100
Navut	C.Santo	100
Lametin	C.Santo	100
Wailapa	S.Santo	100

LANGUAGE	LOCATION	SPEAKERS
Fortsenal	C.Santo	100
Roria	C.Santo	100
Amblong	S.Santo	100
Morouas	C.Santo	100
Tangoa	S.Santo	250
Araki	S.Santo	70
Mafea	E.Santo	50
Tutuba	E.Santo	100
Aore	S.Santo	1
Malo	S.Santo	1500
Narango	S.Santo	160
Tambotalo	S.Santo	50

Malekula Coastal Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Malua Bay	N.W.Malekula	200
Vovo	N.Malekula	100
Mpotovoro	N.Malekula	120
Mae	N.Malekula	250
Vao	N.E.Malekula	900
Atchin	N.E.Malekula	950
Uripiv-Wala-Rano	N.E.Malekula	2300
Unua	S.E.Malekula	300
Rerep	S.E.Malekula	200
Aulua	S.E.Malekula	200
Burmbar	S.E.Malekula	350
Port Sandwich	S.E.Malekula	750
Maskelynes	S.E.Malekula	650
Axamb	S.Malekula	400
Malfaxal	S.W.Malekula	400
South-West Bay	S.W.Malekula	250

Epi Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Lewo	N.Epi	1000
Bierebo	Epi	270
Baki	Epi	100
Maii	Epi	100
Bieria	Epi	70

Central New Hebrides Subgroup:

LANGUAGE	LOCATION	SPEAKERS
Namakura	Shepherd Is.	2000
North Efate	Shepherds/Efate	2500
South Efate	Efate	2200

7. Polynesian Outliers

LANGUAGE	LOCATION	SPEAKERS
Emae	Emae Is.	150
Fila-Mele	S.Efate	1800
Futuna-Aniwa	Futuna/Aniwa Is.	600

The languages of the New Hebrides, then, will be seen to number 105, including the three Polynesian Outlier languages listed under section 7, above. These three languages were deliberately omitted from the classification proper. The language names and breakdowns used in this paper update Tryon 1972. Numbers of speakers, always approximate, have also been updated since that publication, although no official census has been taken for the whole of the New Hebrides since 1967.

2.2. *QUALITATIVE EVIDENCE*

The qualitative evidence for the internal subgrouping of New Hebrides languages has been set out in full in an earlier publication (Tryon 1976:9-59). The major groupings suggested by the phonological evidence are as follows:

(a) All areas of the New Hebrides apart from the Banks and Torres Islands lose POC *R reflecting *daRa(q) *'blood'* and *kuRita *'squid'*.

(b) A large subgroup in the Central New Hebrides (Santo, excepting N.W. Santo, Malekula, West and South-East Ambrym, Paama, Epi, Efate and the Shepherd Islands) would appear to be defined by the loss of *R reflecting *paRi *'stingray'*.

(c) Several subdivisions are suggested within the Central New Hebrides subgroup, the main ones being a grouping which may be called South-Central, consisting of the languages of South-East Ambrym, Paama, Epi, Efate and the Shepherd Islands (based on consonant alternation), and a North-Central subgroup, consisting of the languages of East Santo and North Malekula (based on apico-labial reflexes of POC bilabials).

(d) The southern islands (Tanna, Erromanga and Aneityum) appear to be set apart from the Central and Northern areas, although the evidence which includes Erromanga is rather scanty at this stage.

The phonological evidence, then, suggests that the languages of the New Hebrides may be considered to fall into three subgroups, thus:

I. Northern Subgroup:
Banks and Torres, Maewo, Aoba, North-West Santo, Pentecost and North Ambrym.

II. Central Subgroup:

Santo (except N.W.Santo), Malekula, West and South-East Ambrym, Paama, Epi, Efate and the Shepherd Islands.

III. Southern Subgroup:

Erromanga, Tanna and Aneityum. While the three southern islands would appear separate from the Central and Northern Subgroups on phonological grounds, the evidence adduced in Tryon 1976 is inconclusive as to the homogeneity of the Southern Subgroup. Lynch (this volume) has, however, provided evidence for the separate status of the Southern Subgroup as a valid subgroup within Oceanic languages.

2.3. MORPHOLOGICAL EVIDENCE

Tryon (1973) discussed some selected morphological features of New Hebrides languages, especially noun classification, numeral systems and verb systems and compared the subgroupings reached with sample lexico-statistical groupings. The comparative morphological study, while incomplete and necessarily so because of the current state of knowledge of these languages, suggested the following subgroupings:

(a) Based on Noun Classification:

- a Northern Group, characterised by almost identical categories and noun class markers; a prized possession class, denoted by *pula, has not been found outside this group. The Northern Group includes the Banks and Torres, Maewo, Aoba, north Pentecost, Malo and the larger part of Santo.

- a Central Group, consisting of North-East Santo, North Malekula, Ambrym, Epi and the languages of Tanna in the south. These languages have been observed to have more noun classes than those of the north, together with a number of different categories.

- a Southern Group, consisting of South Malekula, Efate and the Shepherd Islands, Erromanga and Aneityum. This group has only two noun classes for possession.

(b) Based on Numeral Systems:

- the only coherent subgrouping suggested by a study of the numeral systems operating in New Hebrides languages (decimal, imperfect decimal and quinary) appears to be a Southern Group, consisting of the languages of South-East Ambrym, Paama, Epi, Efate and the Shepherd Islands, Erromanga, Tanna and Aneityum. Within these languages a quinary notation is in operation.

(c) Based on Verb Systems:

- an Extreme Northern Group (Banks and Torres Islands)

- a Northern Group (Maewo, Aoba, Pentecost and Santo)

- a Central and Southern Group (the remainder)

The two northern groups form a reasonably homogeneous unit, the main difference between the two being that in the Extreme Northern Group a free form particle system operates, while in the Northern Group the verbal pronoun and tense/aspect markers are fused into a single unit, the morphemes being formally very similar between the two groups. The remaining group, the Central and Southern Group, on the other hand, is characterised by a more complex verb morphology, rather different from that noted in the other two groups. However, the Central and Southern Group cannot be described as homogeneous in terms of verb morphology.

2.4. *CONCLUSIONS AND SYNTHESIS*

As far as the internal relationships of the New Hebrides are concerned, the subgroupings reached through a consideration of the lexicostatistical evidence are in large part confirmed by the evidence from phonology and comparative morphology, although the phonological and morphological evidence has not been as full as might be desired, largely as a result of the lack of data at present available.

On lexicostatistical criteria there are six higher order groupings, as discussed above. Of these six, the three southern islands (Erromanga, Tanna and Aneityum) are seen to be separate from each other and from the other New Hebrides languages. The phonological evidence adduced supports the separation of the southern languages from the languages to the north. The evidence is not sufficient, however, to allow any positive statement concerning the homogeneity or otherwise of the languages of the three southern islands. (Lynch, this volume, suggests that on the basis of reconstructions, the three may be considered to constitute a single subgroup of Oceanic.)

The very large East New Hebrides Subgroup of the North and Central New Hebrides Group, based on lexicostatistics, corresponds very closely to the Northern New Hebrides grouping suggested by the phonological evidence, the only areas at variance being the south-eastern section of Ambrym and Paama.

In other areas the picture produced by comparing the lexicostatistical and phonological evidence is less clear. The phonological evidence produces a rather undifferentiated picture which, although not conflicting with the lexicostatistical picture, fails to support as detailed a subgrouping as the quantitative evidence requires.

The morphological evidence produced to date does very little except to confirm the north-eastern grouping suggested by the lexicostatistical and phonological evidence. For the remainder it is inconclusive. While comparative morphological studies are proceeding, the present state of knowledge is insufficient to allow a detailed morphologically-based classification. For the time being, at least, the internal New Hebrides picture will have to rely upon mainly lexical and phonological data, as suggested above.

3.0. EXTERNAL RELATIONSHIPS

3.1. *SELECTING REPRESENTATIVE LANGUAGES*

The Lexicostatistical classification of the languages of the New Hebrides revealed that internally at least there were six major groupings, as follows:

1. North and Central New Hebrides Group
2. East Santo Group
3. Malekula Interior Group
4. Erromanga Group
5. Tanna Group
6. Aneityum Group

The following languages were chosen to represent each of the six groups:

Group 1: Mota and North-East Aoban
Group 2: Sakao
Group 3: Big Nambas
Group 4: Sie
Group 5: Lenakel
Group 6: Aneityum

Because of the size of the Group, two languages were chosen to represent Group 1, Mota and North-East Aoban.

In order to attempt to place the languages of the New Hebrides within the total framework of the Austronesian languages of the Pacific, it was decided to limit the study to exploring the relationships of the New Hebrides languages with the so-called Oceanic languages, or Eastern Austronesian. While it had been hoped to consider phonological and morphological information, time and data available have limited the writer to a lexicostatistical computation of the external relationships of New Hebrides languages only, at this stage.

As a basis for selecting the languages to be compared with the New Hebrides representatives, reference was made to Grace (1955), who made the first attempt at an all-embracing subgrouping of Oceanic languages.

In his report, based on 400 Oceanic languages, and taking into account
both lexical and morphological information, Grace produced a tentative
classification which resulted in his classifying all Oceanic languages
into 19 subgroups (Grace 1955:338-9), as follows:

1.   New Caledonia

2.   Nengone (Loyalty Islands)

3.   Lifu (Loyalty Islands)

4.   Iai (Loyalty Islands)

5a. Southern New Hebrides (Aneityum, Tanna, Erromanga)

5b. Remainder of New Hebrides except North-East New Hebrides/
     Rotuma/Fiji/Polynesian

5c. Pentecost/Maewo/Aoba/Banks/Torres Is

5d. Micronesian (?)

6.   Santa Cruz (?)

7.   South-eastern Solomons

8.   New Georgia Archipelago

9.   Choiseul

10. Bougainville Straits, Bougainville, Buka

11. New Ireland, Hanover, Duke of York, North New Britain

12. South-West New Britain, Kobe, French Is, Siassi Is, Kelana,
     Tami, Yabim, Bukaua, Suam

13. Astrolabe Bay Area

14. Manam and Schouten Is

15. Remainder of Sepik District

16. Admiralty Is and Western Is (except Wuvulu and Aua)

17. Wuvulu and Aua

18. Central District of Papua

19. Milne Bay and Northern Districts of Papua.

Pawley (1972:5) reports that Grace has made some alterations to his 1955
grouping, and in a more recent paper (Pawley forthcoming:10) reports yet
another minor modification based on an unpublished paper (Grace 1971).
The present writer has preferred to select languages representative of
Grace's 1955 grouping for comparison with representatives of the six
major New Hebrides groupings established (Tryon 1976), largely because
of materials available. The languages chosen for comparison with the
New Hebrides languages were as follows:

1. Xaraci (New Caledonia)

2. Nengone (Loyalties) (Grace's original 2/3/4, subsequently merged)

3. Gilbertese and Samoan (Grace Group 5)

4. Kwaio (South-east Solomons) (Grace Group 7; Santa Cruz omitted)

5. Roviana (Grace Group 8, New Georgia)

    6. Sengga (Choiseul, Grace Group 9)

    7. Halia (Grace Group 10)

    8. Nakanai (Grace Group 11)

    9. Yabem (Grace Group 12)

  10. Gedaged (Grace Group 13)

  11. Kairiru (Grace Group 14)

  12. Ali (Grace Group 15, Sepik)

  13. Titan (Grace Groups 16/17; combined following Blust personal
       communication)

  14. Motu (Grace Group 18)

  15. Kiriwina (Grace Group 19)[1]

The above 16 non-New Hebrides languages, together with the seven New
Hebrides languages selected as representatives gave a total of 23 lan-
guages compared. For interest's sake, two languages were added, namely
Mor (Irian Jaya) and Toba Batak (Indonesia);[1] the percentages for these
last two were computed but excluded from the subgrouping proposed, as
they fell outside the original Oceanic area as set up by Grace in 1955.

3.2. *THE QUANTITATIVE EVIDENCE*
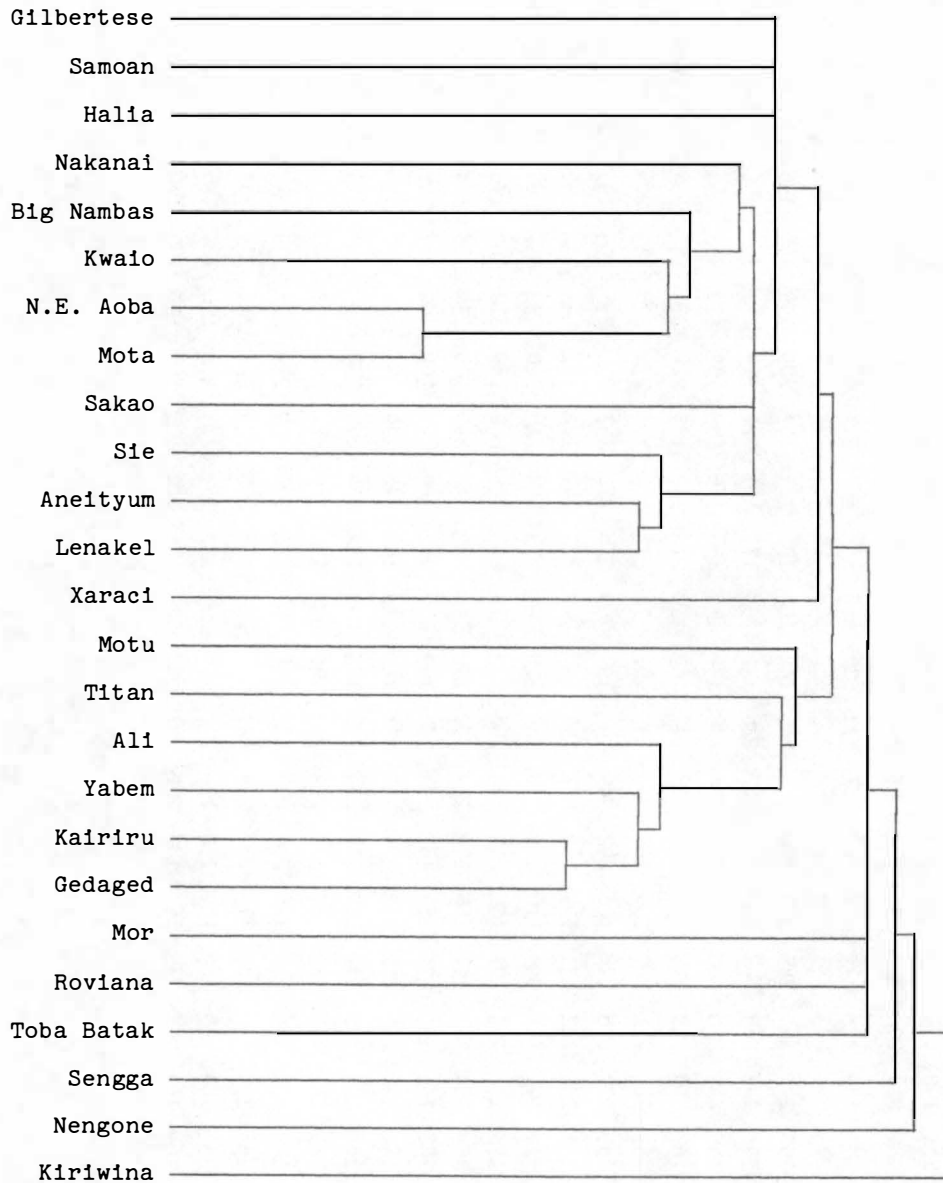
    In order to make the computation on which the external relationships
of the New Hebrides languages is based, the wordlists used contained
exactly the same items as those used for the calculation of the internal
relationships (see Tryon 1976:67ff). A maximum of 258 items were com-
pared, with a minimum of around 200. The percentages of shared cognates
computed is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
–	206	183	206	188	216	174	188	181	223	241	209	130	084	157	162	159	193	194	188	189	179	133	184	146	XAR
206	–	113	124	100	128	098	111	097	106	122	116	074	037	072	097	076	102	100	079	081	109	084	108	103	NEN
183	113	–	257	161	190	174	179	167	229	226	216	142	119	193	208	158	192	207	174	164	211	154	205	167	GIL
206	124	257	–	180	228	202	212	196	237	257	247	167	130	187	221	169	215	209	200	213	231	168	218	166	SAM
188	100	161	180	–	243	150	131	145	267	289	199	142	120	175	169	188	182	168	161	139	158	134	161	166	SAK
216	128	190	228	243	–	243	243	240	299	289	223	152	121	176	184	191	196	184	163	191	198	161	182	200	BIG.N
174	098	174	202	150	243	–	285	270	240	211	206	142	107	175	162	167	203	195	165	156	174	159	170	194	SIE
188	111	179	212	131	243	285	–	305	219	207	228	131	099	177	191	171	214	212	183	186	187	180	175	210	LEN
181	097	167	196	145	240	270	305	–	240	226	208	134	110	158	166	163	183	170	187	186	175	130	157	175	ANE
223	106	229	237	267	299	240	219	240	–	468	242	169	120	212	213	185	224	214	185	204	201	158	187	174	MOT
241	122	226	257	289	289	211	207	226	468	–	303	201	170	243	254	198	232	228	213	209	214	154	222	214	NEA
209	116	216	247	199	223	206	228	208	242	303	–	194	169	258	272	197	231	235	203	195	224	157	207	192	KWA
130	074	142	167	142	152	142	131	134	169	201	194	–	210	223	142	143	141	143	130	136	161	109	157	155	ROV
084	037	119	130	120	121	107	099	110	120	170	169	210	–	179	142	116	120	099	097	088	126	109	120	104	SEN
157	072	193	187	175	176	175	177	158	212	243	258	223	179	–	209	202	189	212	164	170	168	130	185	208	HAL
162	097	208	221	169	184	162	191	166	213	254	272	142	142	209	–	209	226	234	188	164	225	162	202	164	NAK
159	076	158	169	188	191	167	171	163	185	198	197	143	116	202	209	–	309	249	217	200	179	173	171	188	YAB
193	102	192	215	182	196	203	214	183	224	232	231	141	120	189	226	309	–	350	272	239	227	198	186	211	GED
194	100	207	209	168	184	195	212	170	214	228	235	143	099	212	234	249	350	–	301	215	245	190	190	223	KAI
188	079	174	200	161	163	165	183	187	185	213	203	130	097	164	188	217	272	301	–	249	225	189	185	215	ALI
189	081	164	213	139	191	156	186	186	204	209	195	136	088	170	164	200	239	215	249	–	207	148	176	193	TIT
179	109	211	231	158	198	174	187	175	201	214	224	161	126	168	225	179	227	245	225	207	–	187	204	219	MOT
133	084	154	168	134	161	159	180	130	158	154	157	109	109	130	162	173	198	190	189	148	187	–	150	183	KIR
184	108	205	218	161	182	170	175	157	187	222	207	157	120	185	202	171	186	190	185	176	204	150	–	207	TOB
146	103	167	166	166	200	194	210	175	174	214	192	155	104	208	164	188	211	223	215	193	219	183	207	–	MOR

Key to abbreviations: 1. Xaraci; 2. Nengone; 3. Gilbertese; 4. Samoan; 5. Sakao; 6. Big Nambas; 7. Sie; 8. Lenakel;
9. Aneityum; 10. Mota; 11. North-East Aoba; 12. Kwaio; 13. Roviana; 14. Sengga; 15. Halia;
16. Nakanai; 17. Yabem; 18. Gedaged; 19. Kairiru; 20. Ali; 21. Titan; 22. Motu; 23. Kiriwina;
24. Toba Batak; 25. Mor

Percentages given to one decimal place; thus 219 = 21.9%

CHART II

```
Gilbertese ──────────────────────────────────────────┐
                                                       │
Samoan ─────────────────────────────────────────────┐ │
                                                     │ │
Halia ──────────────────────────────────────────────┤ │
                                                     ├─┤
Nakanai ──────────────────────────────────┐          │ │
                                           │          │ │
Big Nambas ──────────────────────────────┐ │          │ │
                                         ├─┤          │ │
Kwaio ───────────────────────────────┐   │ │          │ │
                                     ├───┘ │          │ │
N.E. Aoba ──────────────────┐         │     │          │ │
                            ├─────────┘     │          │ │
Mota ───────────────────────┘               │          │ │
                                            ├──────────┘ │
Sakao ──────────────────────────────────────┤            │
                                            │            │
Sie ─────────────────────────────┐          │            │
                                 ├──┐        │            │
Aneityum ────────────────────────┘  │        │            │
                                    ├───────┤            │
Lenakel ────────────────────────────┘        │            │
                                             ├───────────┤
Xaraci ──────────────────────────────────────┘            │
                                                          │
Motu ────────────────────────────────────┐                │
                                          ├──┐             │
Titan ───────────────────────────────────┘  │             │
                                             ├──┐          │
Ali ──────────────────────────────────────┐  │  │          │
                                          ├──┘  │          │
Yabem ────────────────────────────────────┘     │          │
                                                ├─────────┐│
Kairiru ─────────────────────────┐               │         ││
                                 ├───────────────┘         ││
Gedaged ─────────────────────────┘                         ├┤
                                                           ││
Mor ───────────────────────────────────────────────────┐  ││
                                                        ├──┘│
Roviana ────────────────────────────────────────────────┘   │
                                                            │
Toba Batak ──────────────┐                                   │
                         ├──────────────────────────────────┤
Sengga ──────────────────┘                                   │
                                                            │
Nengone ─────────────────────────────────────────────────────┘
                                                            │
Kiriwina ────────────────────────────────────────────────────┘
```

3.3. *TWO INTERPRETATIONS*

The percentage figures have been arranged in the traditional tree form in Chart II above. They show, first of all, that the six higher order groups established for the internal New Hebrides classification hold good even when they are compared with other Oceanic groups. In order that a non-New Hebridean language should be a member of one of the six New Hebrides groups established (Tryon 1976), it would have to share more than 30% cognates. The only case in which this condition is fulfilled is that of Kwaio, in the South-East Solomons and North-East Aoba, which share 30.3% according to the computation carried out. In this case, the languages of the South-East Solomons could tentatively be included in the East New Hebrides Subgroup, as represented in Chart I, since Kwaio was chosen to represent the languages of that Grace-established subgroup.

The percentages computed, see the table above, tend to fall into a uniform range, something between 10 and a little over 20%, resulting in a rather undifferentiated subgrouping picture, as Chart II shows. It was found that until the percentages reached roughly 25% or higher, it was not possible to achieve any real differentiation. For this reason, Chart III, based on cognate densities of higher than 25%, was produced, showing a differentiated grouping which suggests the following:

(a) The existence of a New Guinea subgroup, within Oceanic languages.

(b) The existence of a South New Hebrides subgroup.

(c) The existence of a subgroup which takes in the remainder of the language groups of the New Hebrides, together with representatives of the languages of the South-East Solomons, together with Grace's Groups 10 and 11, and Micronesian (as represented by Gilbertese).

(d) That the remaining groups isolated by Grace do not have affiliations of the same order as a, b and c, above.

The above subgroupings must, of necessity, be regarded only as very tentative, since they would require corroborative evidence from com-parative morphology and phonology. The New Guinea subgroup has in fact been supported by Pawley (forthcoming), and Lynch (this volume) provides evidence for a higher order grouping which includes the three southern islands of the New Hebrides.

Apart from the fact that they produce a rather undifferentiated picture, cognate densities in the very low range, say less than 20-25%, must be regarded as suspect, and consequently unreliable as a means of subgrouping, for at such low percentages the possibility of not

CHART III
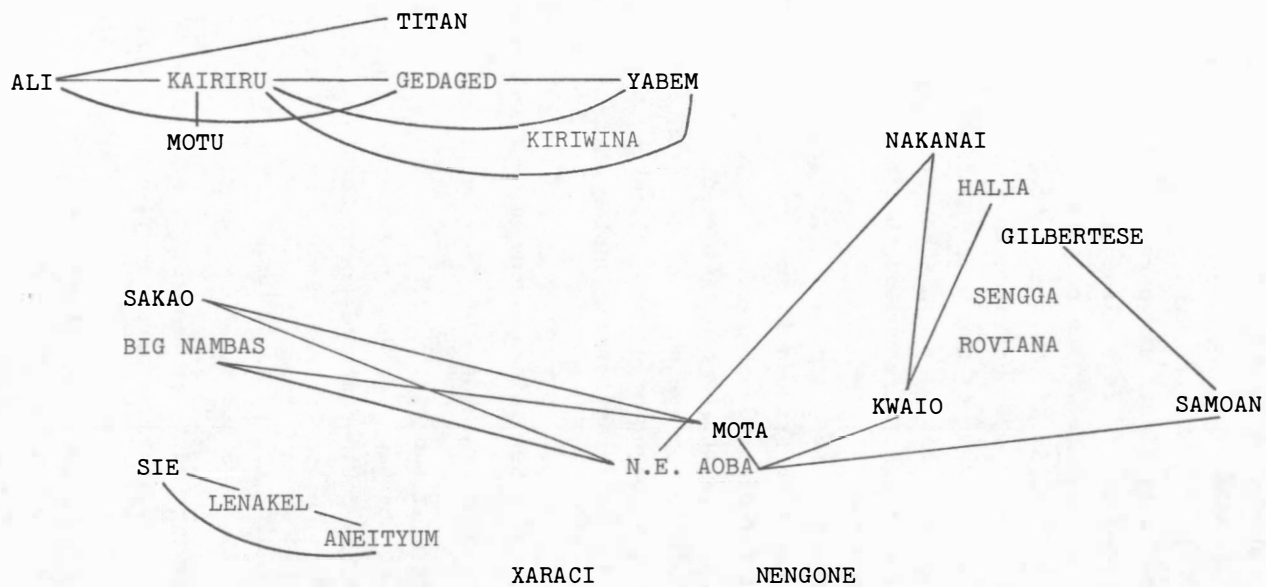
recognising cognates becomes much higher than, say, in the range 60-70%, where cognate recognition poses far fewer problems. Unless all sound correspondences have been determined, especially at this level, the number of doubtful cognates must increase in inverse proportion to the cognate percentages. Mathematically, too, the problem of standard error and degree of confidence in the correctness of recognitions becomes more pressing as the cognate densities become lower. In point of fact, it is suggested that where percentages of less than say 20-25% are involved, lexicostatistics is at best a most unsatisfactory means of subgrouping. At this level the evidence of phonology, comparative morphology and lexical reconstruction must be preferred.

When low percentages are involved, Guy (forthcoming) has proposed that a new technique involving linear regression might well be applied and prove more worthwhile than the lexicostatistical techniques that have become traditional. The technique has been applied to the percentages computed for the languages considered in this paper, and appears as Chart IV. The clusterings produced may be described as an index of predictability, rather than a genetic subgrouping, for what the method achieves is to allow one, knowing the relationship that exists between two or more languages, to determine which language will best predict the next in the string to be considered. The technique is at the experimental stage only, but the clusterings produced are worth considering, especially when compared to those of Chart III. Linear regression techniques have been applied successfully to other disciplines, and may well have a useful role to fill in linguistic theory as well.

4.0. CONCLUSIONS

The languages of the New Hebrides may be considered to fall into six higher order subgroups in terms of internal classification, based on quantitative evidence and supported to a large extent by qualitative evidence.

In terms of external relationships, the findings presented here must be regarded as an approach to the question rather than a definitive subgrouping, for, apart from the fact that they are based on quantitative evidence alone, the present state of knowledge of the Austronesian languages of Melanesia is such that it must preclude the possibility of anything like definitive statements. Other groupings, based on lexicostatistics, have been produced in the past, but they too have been thwarted by the problem of uncertainty occasioned by low cognate densities.

CHART IV

Clustering from Percentage Correlation Coefficients (Mor and Toba Batak Excluded)
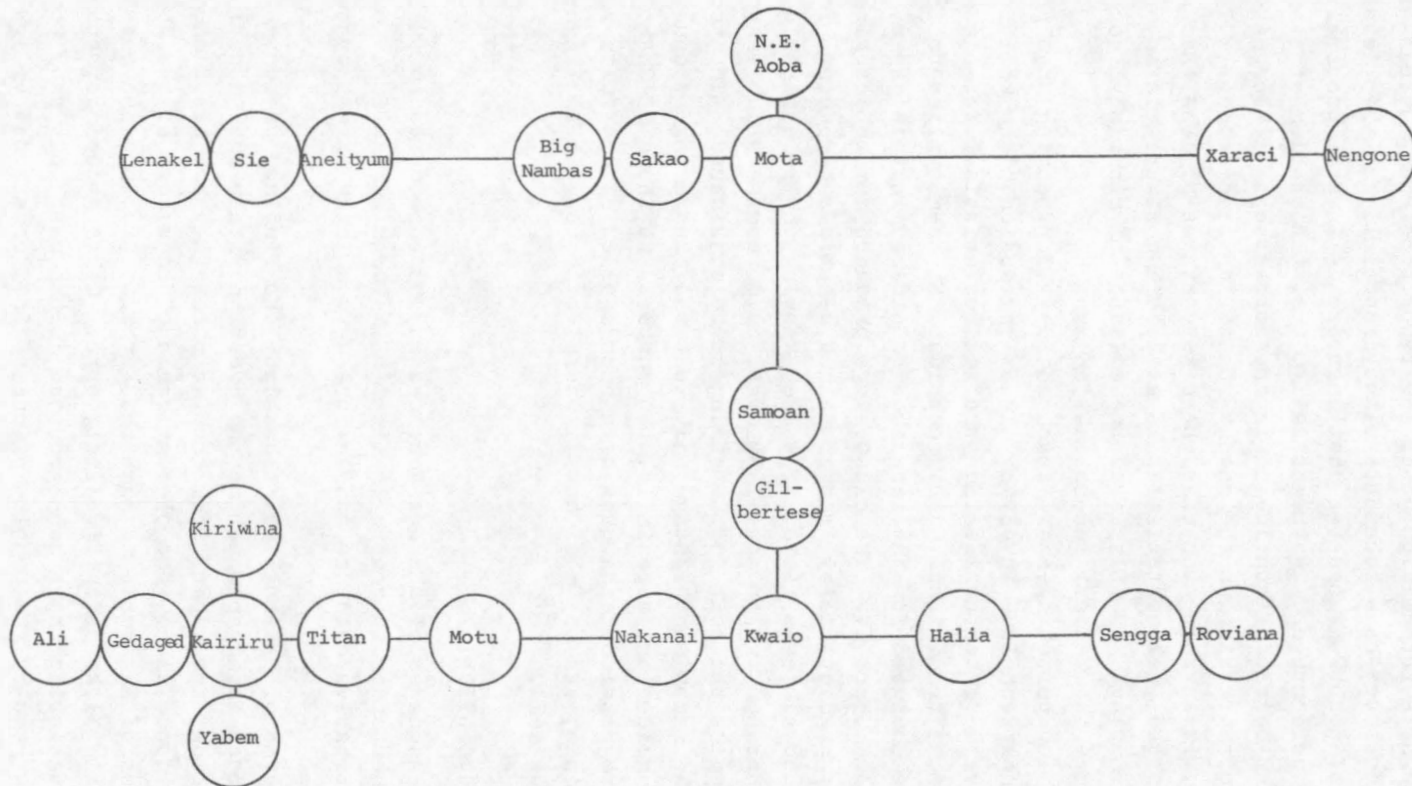
Distance between the centre of any two circles linked by a
line is inversely proportional to r^5 $(d_{cm} = \frac{1}{r^5})$

As regards the external relationships of the New Hebrides languages, the best that may be said at present is that on lexicostatistical grounds the southern languages may be considered to constitute a sub-group within the Oceanic framework, while the remainder of the languages of the New Hebrides together with those of the South-East Solomons, the Gilberts and the general Bougainville area (Grace's Groups 10 and 11), constitute another large Oceanic subgroup.

There is, however, a considerable amount of research to be done before anything but a tentative statement of relationships may be made, for the evidence of comparative phonology and morphology as well as lexical reconstruction must be given careful consideration. For while data for quantitative studies is now reasonably widely available, such studies must be regarded as only a first step, requiring confirmation or modification as more detailed and systematic information becomes available. Detailed morphological studies are in progress in many areas of the New Hebrides, but given the large number of languages both there and in other areas of Melanesia, one might expect that statements concerning relationships will remain tentative for some little time.

N O T E

1. Thanks are due to R.S. Lawton, J.P. Sarumpaet, R.L. Johnston, D.C. Laycock, J.F. Streicher, R.A. Blust and others whose materials assisted greatly in the compilation of data.

BIBLIOGRAPHY

ALLEN, J. and J. ALLEN

1965   *Halia Language Course*. Port Moresby: Department of
        Information and Extension Services.

BLUST, R.A.

forth-   'The Proto-Oceanic Palatals'. To appear as Memoir 24 of
coming   the Polynesian Society, Auckland.

n.d.   Titan Wordlist.   Typescript.

CAPELL, A.

1954   *A Linguistic Survey of the South-Western Pacific*. Noumea:
        South Pacific Commission, Technical Paper 70. (Revised
        1962.)

CHARPENTIER, J.M.

1974   Langue de Port Sandwich.   Thèse de Doctorat de Troisième
        Cycle, Université de Bordeaux III.

CHATTERTON, Percy

n.d.   *A Basic Motu Dictionary*, bound with *A Grammar of the Motu
        Language of Papua*, by R. Lister-Turner and J.B. Clark,
        second edition, edited by P. Chatterton. Sydney: Government
        Printer.

CODRINGTON, R.H.

1885   *The Melanesian Languages*. Oxford: Clarendon Press.

DYEN, I.
   1965   *A Lexicostatistical Classification of the Austronesian Languages*. *IUPAL* 19. *IJAL* Memoir 19.

GABELENTZ, H.C. von der
   1861-73 *Die melanesischen Sprachen* ... Abhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften, philologisch-historische Klasse 3 & 7. Leipzig.

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

   1969   'A Proto-Oceanic Finder List'. *WPLUH* 1/2:39-84.

   1971   Tentative Oceanic Subgrouping. University of Hawaii. Typescript (not for circulation).

   1975   *Canala Dictionary (New Caledonia)*. PL, C-2.

GUY, J.B.M.
   1974   *A Grammar of the Northern Dialect of Sakao*. PL, B-33.

   forth- 'Linear Regression and Subgrouping in Linguistics'.
   coming

HACKMAN, B.D.
   n.d.   Sengga Wordlist.

HEWITT, H.J.
   1966   'Aneityum of the Southern New Hebrides: Anejom Segmental Phonology and Word List: A Preliminary Report'. *Te Reo* 9:1-43.

JOHNSTON, R.L.
   n.d.   Nakanai Wordlist.

KASARHÉROU, J.
   1962   'Les changements vocaliques de trois préfixes en motlav'.
             Te Reo 5:32-4.

KEESING, R.M.
   1975   *Kwaio Dictionary.* PL, C-35.

KERN, H.
   1906   *Taalvergelijkende verhandeling over het Aneityumsch. Met
             een Aanhangsel over het klankstelsel van het Eromanga.*
             VKNA 8/2.

LAWTON, Ralph S.
   forth-   Kiriwina Dictionary. MS. To appear in *PL*.
   coming

LAYCOCK, D.C.
   n.d.a   Ali Wordlist. MS.

   n.d.b   Kairiru Wordlist. MS.

   n.d.c   Mor Wordlist. MS.

LYNCH, J.D.
   1974   Lenakel Phonology. Ph.D. dissertation, University of
             Hawaii.

   1975   The Southern New Hebrides and Proto-Oceanic: A Report on
             Work in Progress. Paper presented to the 9th Annual
             Congress of the Linguistic Society of Papua New Guinea,
             Goroka.

   1976   Summary of Southern New Hebrides Historical Phonology.
             Typescript.

   1978   'Proto-South Hebridean and Proto-Oceanic'. See pp.717-79
             in this volume.

MAGER, J.F.
   1952   *Gedaged-English Dictionary.* Columbus, Ohio: Board of
             Foreign Missions of the American Lutheran Church.

PARKER, G.J.
    1968   'Southeast Ambrym Verb Inflection and Morphophonemics'.
            PL, A-15:27-40.

    1970   *Southeast Ambrym Dictionary*. *PL*, C-17.

PATON, W.F.
    1971   *Ambrym (Lonwolwol) Grammar*. *PL*, B-19.

    1973   *Ambrym (Lonwolwol) Dictionary*. *PL*, C-21.

PAWLEY, A.K.
    1972   'On the Internal Relationships of Eastern Oceanic
            Languages'. In: R.C. Green and M. Kelly, eds *Studies in*
            Oceanic Culture History 3:1-142. Pacific Anthropological
            Records 13. Honolulu: Bishop Museum.

    forth-   The New Guinea Oceanic Hypothesis. To appear in *WPLUH*.
    coming

RAY, S.H.
    1926   *A Comparative Study of the Melanesian Island Languages*.
            Cambridge University Press.

SABATIER, E.
    1971   *Gilbertese-English Dictionary*. Sydney: South Pacific
            Commission Publications Bureau.

SAMARIN, W.J.
    1967   *Field Linguistics*. New York: Holt, Rinehart & Winston.

SCHÜTZ, A.J.
    1969a   *Nguna Texts*. *OL* Special Publication No.4. Honolulu:
            University of Hawaii Press.

    1969b   *Nguna Grammar*. *OL* Special Publication No.5. Honolulu:
            University of Hawaii Press.

STREICHER, J.F.
    1976   Yabem Dictionary. MS.

TRYON, D.T.

    1972   'The Languages of the New Hebrides: A Checklist and
              General Survey'. *PL*, A-35:43-84.

    1973   'Linguistic Subgrouping in the New Hebrides: a Preliminary
              Approach'. *OL* 12:303-51.

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

TRYON, D.T. and M.-J. DUBOIS
    1969   *Nengone Dictionary. Part I: Nengone-English*. PL, C-9.

WALSH, D.S.
    1962   'The Phonemes of Raga'. *Te Reo* 5:57-60.

    1966   The Phonology and Phrase Structure of Raxa. M.A. thesis,
              University of Auckland.

WATERHOUSE, J.H.L.
    1949   *A Roviana and English Dictionary* ... revised by L.M. Jones;
              Sydney: Epworth Press.