Chapter 2. Proto-Austronesian and the Major Austronesian Subgroups

Darrell Tryon

The discovery of the existence of the Austronesian language family goes back to the seventeenth century, when members of the Schouten and Lemaire expedition collected a vocabulary from East Futuna (Wallis and Futuna) in the South Pacific, some of which was found later to bear a striking resemblance to Malay. However, it was not until the nineteenth century that Austronesian studies really took on a systematic flavour, particularly with the work of Dutch linguists in what is today Indonesia and that of a number of missionaries in the South Pacific. A discussion of the history of research in Austronesian linguistics is not called for here. Suffice it to say that the major advances in Austronesian studies have been made this century, beginning with the systematic comparative work of such linguists as Stresemann (1927) and Dempwolff (1934-38). Since that time there has been a great deal of systematic research carried out right throughout the vast region where Austronesian languages are spoken.

In the past twenty-five years or so there have been a number of subgrouping hypotheses advanced by scholars of Austronesian languages. Only the most recent will be considered in any detail in this overview, for the major purpose of this paper is to present current and recent Austronesian subgrouping hypotheses, to look at what may be regarded as secure and what remains the subject of ongoing research. But before undertaking this review, it is necessary to insert a couple of preliminary remarks, one on the make-up of the Austronesian language family and the other on methodological considerations, in particular the methodology of linguistic subgrouping.

The Austronesian Family

The Austronesian language family is perhaps the world's largest, with some 1200 languages and approximately 270 million speakers, according to the most recent studies (Tryon, ed. 1994). It ranges from languages with tens of millions of speakers (Malay/Indonesian, Javanese, Tagalog) to a surprisingly large number of languages with only a handful of speakers, numbered in the hundreds. These latter are particularly prevalent in Oceania, the causes for which will be examined later in this volume by Dutton. The geographical range of the Austronesian family is displayed in Map 1. A glance at the map will show that Austronesian languages are spoken from Madagascar in the west to Easter Island in the east. They are spoken almost universally in Indonesia and the Philippines, in Singapore and Malaysia, by the indigenous population of Taiwan, and by minority

populations in Vietnam, Cambodia and the Mergui Archipelago off the coast of Burma (currently Myanmar). Further east, Austronesian languages occupy almost all of the islands of Oceania with the exception of the inland and most coastal areas of the great island of New Guinea (Irian Jaya and Papua New Guinea).
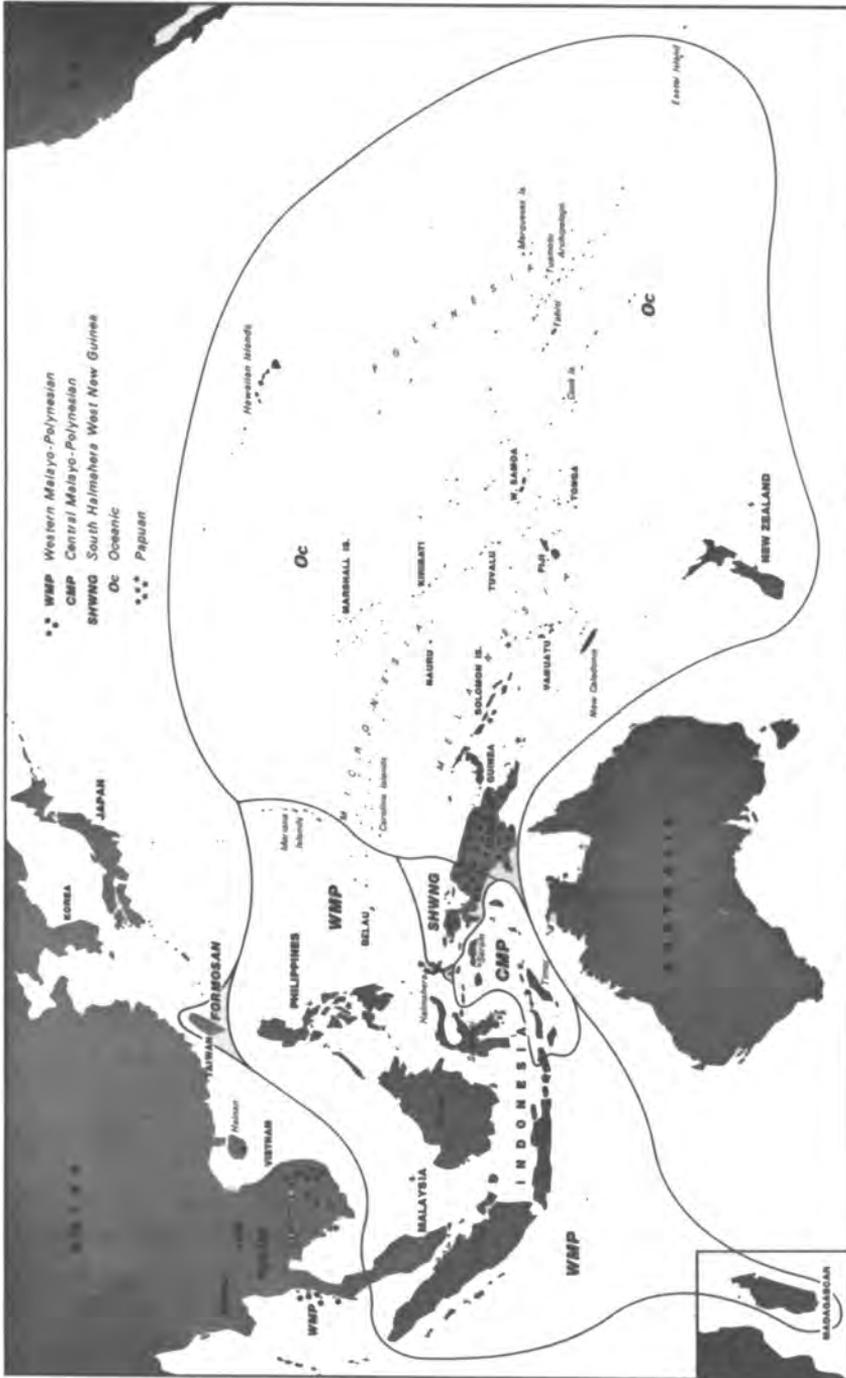
**Map 1: The geographical range of the Austronesian family.**

Methodology

The principal method that has been used to subgroup the Austronesian languages has been the traditional comparative-historical method, largely developed last century in connection with the comparative study of the Indo-European languages. Very briefly, this method systematically compares the regular sound correspondences between the languages compared as a first step towards reconstructing the proto-language from which the daughter languages have descended. Once the reconstruction of the proto-language has been achieved, then individual languages and sets of languages are examined to determine the innovations (phonological, morpho-syntactic and lexical) which they reflect relative to the proto-language. It is upon these innovations that subgrouping depends and proceeds.

Ross (1994) has made two very pertinent observations concerning the way in which these innovations are distributed across languages. He notes that innovations pattern across languages in two different ways reflecting two different developmental sequences. In the first, groups of languages share discrete bundles of innovations. Thus taking, for example languages A to Z, members of the family whose proto-language or putative ancestor language is *AZ, languages A-P may share one bundle of innovations not shared by languages Q-Z and vice versa. These languages would then fall into two distinct subgroups, AP and QZ. This is the distribution of innovations which results when languages have diversified by separation, that is when two or more communities speaking the same language become geographically separated. However, this is not the only way in which languages diversify. They also diversify without physical separation through dialect differentiation in their home territory. In this situation, instead of discrete bundles of innovations there are overlapping bundles which form a chain. Both of these distributional phenomena are observable in the Austronesian family, and will be seen to have important consequences for Austronesian subgrouping today. While the comparative method is a powerful tool, it has some limitations, especially with problems of recognizing contact-induced language change (see the chapter by Tom Dutton in this volume).
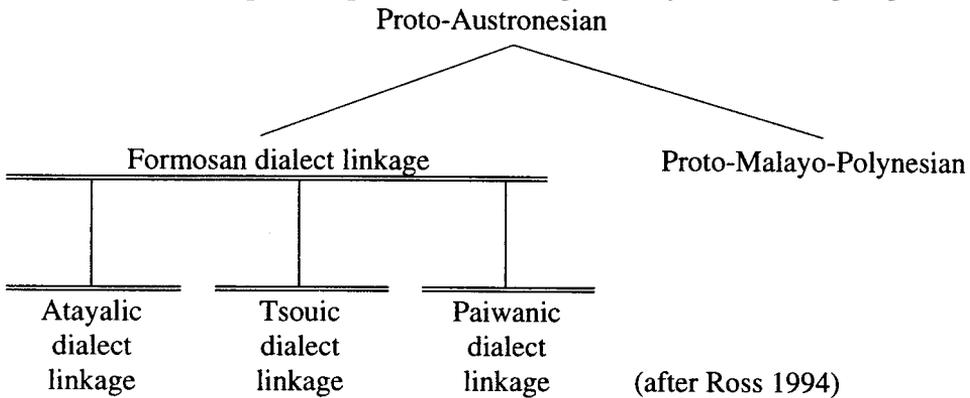
Another method used to subgroup languages is lexicostatistics, a method based on the replacement rate of the basic lexicon of a language over time. This method was employed by Dyen (1965) in his well-known lexicostatistical classification of the Austronesian family. While this method is useful as a first approximation, it is most useful with languages which are quite closely related. The major problem with this method, however, is that it is based on a premise that all languages replace vocabulary at a constant rate, which claim is demonstrably erroneous. For this reason the subgrouping hypotheses discussed in this paper are all based on comparative-historical techniques.

Current Subgrouping Hypotheses

Proto-Austronesian, the ancestral language from which all other Austronesian languages descended, is considered by most scholars to have been spoken on the island of Taiwan something in the order of 5000 years ago. This ancestral language is considered to have diverged over time into four major subgroups, represented as follows:

Proto-Austronesian

Atayalic   Tsouic   Paiwanic   Malayo-Polynesian
(after Blust 1977)
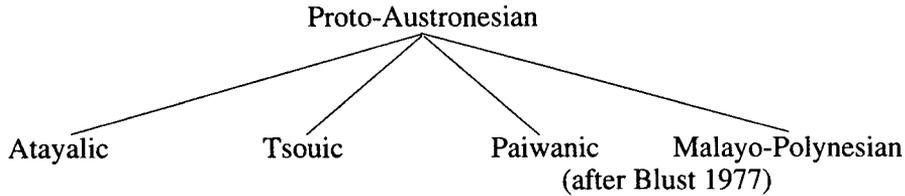
In other words many scholars consider that three of the four highest-order subgroups of Austronesian are spoken on Taiwan and have been ever since the development of Proto-Austronesian. As discussed above, languages evolve by two different processes — gradual dialect differentiation, and separation. It is likely that the languages of the Atayalic, Tsouic and Paiwanic subgroups have arisen by gradual dialect differentiation from Proto-Austronesian, or from early descendant dialects spoken by the population which stayed behind when the languages which belong to the Malayo-Polynesian subgroup left the island. This distinction in developmental process is better signalled by the following diagram:

Proto-Austronesian

Formosan dialect linkage   Proto-Malayo-Polynesian

Atayalic   Tsouic   Paiwanic
dialect   dialect   dialect
linkage   linkage   linkage   (after Ross 1994)

According to this classification all of the Austronesian languages spoken outside Taiwan are descended from Proto-Malayo-Polynesian.

At this point it would be useful to consider Blust's representation of all of the major subgroups of Austronesian and then to return and consider the evidence upon which they are based, together with some alternative subgroupings currently under consideration by Austronesianists. The full Austronesian family tree devised by Blust (1978) is as follows:
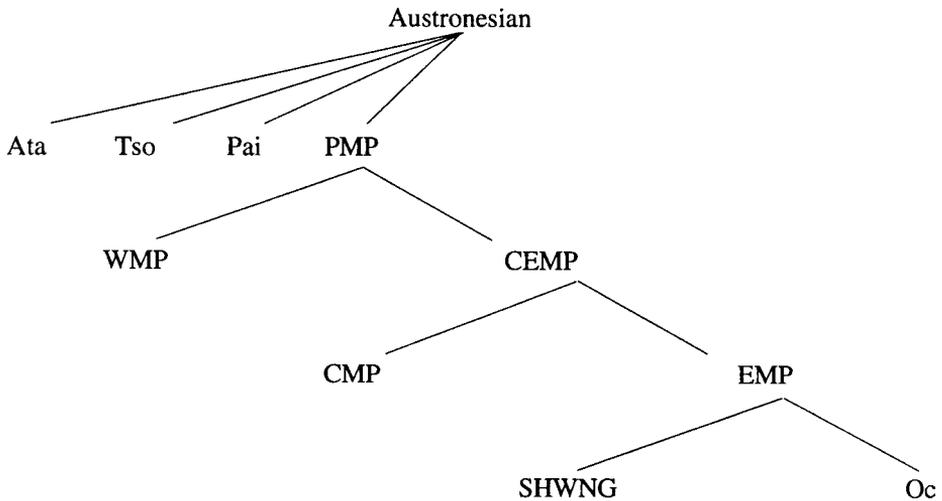
Ata	Atayalic	CEMP	Central-Eastern Malayo-Polynesian
Tso	Tsouic	CMP	Central Malayo-Polynesian
Pai	Paiwanic	EMP	Eastern Malayo-Polynesian
PMP	Proto-Malayo-Polynesian	SHWNG	South Halmahera-West New Guinea
WMP	Western Malayo-Polynesian	Oc	Oceanic

Perhaps a good place to begin is to examine the major pieces of evidence which led scholars to conclude that all Austronesian languages outside Taiwan constitute a single first-order subgroup of Austronesian. Blust (1977, 1982) adduces the following:

1. The PAn pronouns fall into two partially distinct sets, a long form (actor/patient) and a short form (agent/possessor). Short forms typically consist of the last vowel plus any preceding consonant of the corresponding long form (e.g. *aku: ku, '1st sg'). Internal Formosan comparison indicates an ancestral system in which this formal correspondence is regular. However, in all regions outside Formosa we find that the short form of the pronoun corresponding to *kamu '2nd pl' is typically, although not exclusively used as a singular pronoun, explained as a "politeness shift". Blust considers that symmetry and simplicity favour treating this usage as an innovation, and thus a valuable piece of subgrouping evidence. Given the unnaturalness of a plural-to-singular shift which affects the short form pronoun and not the corresponding long form, it is simplest to attribute the singular use of *-mu to one change rather than to a series of convergent innovations. Blust concludes that the change *-mu '2nd pl' > -mu '2nd sg' is therefore taken as evidence for a non-Formosan (Malayo-Polynesian) subgroup of the Austronesian languages (1982:235).

2. In an earlier paper, Blust (1977:11-12) cites a further piece of pronominal evidence supporting his claim. This concerns long forms or full forms. Blust states that Dahl had reconstructed a form *a(N)ken '1st sg full form' in addition to *aku. Blust suggests that in languages outside Taiwan the *-en suffix indicates goal focus, while on the Proto-Austronesian level the only meaning that can be securely attributed to *a(N)ken is that of absolute possession, citing also PAn *iten v *kita '1st pl.inc.', and *amen v *kami '1st pl.exc.'.

3. A phonological corroboration of the subgrouping of the Austronesian languages outside Taiwan into a single higher-order subgroup is the following: PAn preconsonantal and final *S disappears, the resultant final shewa merging with *a in all Malayo-Polynesian languages. Thus, for example:

> PAn *kuSkuS > PMP *kuku 'nail (of finger, toe)',
> PAn *tuqaS > PMP *tuqa 'old',
> PAn *CumeS > PMP *tuma 'clothes louse'.

4. At the same time, another innovation which is reflected right across the Malayo-Polynesian region is the use of the PMP verbal prefixes *pa ŋ-, and *maŋ- to form verbs where the agent is the subject from verbs where the patient is subject (Ross 1994; Reid, pers.comm.).

5. In addition to the above evidence, there are a number of other phonological innovations upon which the PMP subgroup is based. These include the merger of PAn *t and *ts as PMP *t. It should be noted, however, that of the languages of Taiwan both Amis and Bunun share this phonological merger. They do, however, differentiate from PMP in terms of criteria 1-4, above.

Dyen (1990) dissents from the view that all of the Austronesian languages outside Taiwan are members of a single Malayo-Polynesian subgroup. Invoking a lexical method called "homomeric lexical classification" whereby "different sets of cognates distributed over exactly the same set of languages are said to be homomerous" (1990:212), Dyen claims that "all the other classifications separate the Philippine languages from the Formosan at, or nearly at, the highest level, whereas the evidence presented here favors regarding the Philippine languages as the closest relatives of the Formosan languages, the latter being considered to form a single subgroup" (1990:224).
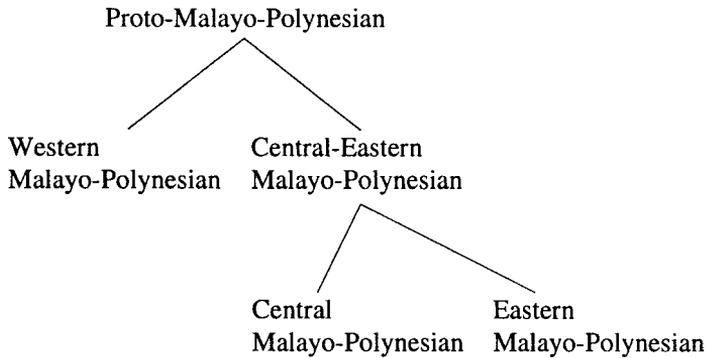
In his discussion of problems in Austronesian subgrouping, Ross (1994) assesses that within Taiwan there is a fair measure of agreement concerning the lower-order subgroups. Li (1980, 1981, 1985) has carried out detailed comparative work on the Atayalic subgroup, and Tsuchida (1976) has produced a substantial reconstruction of Proto-Tsouic. There is also general agreement on the core

members of the Paiwanic subgroup. Beyond this there are disagreements as to subgroup affiliation, especially regarding the position of Rukai. Compare, for instance, the family trees produced in Tsuchida (1976) and Li (1985). Li (1985) proposes three major subgroups within Taiwan; a Northern group, which includes a number of languages attributed by others to Paiwanic, Tsouic, and a reduced Paiwanic group. In spite of subgrouping problems with the Austronesian languages of Taiwan, it appears clear that Proto-Austronesian diversified into a linkage of dialects and/or languages before the speakers of what later became Proto-Malayo-Polynesian (PMP) left Taiwan.

Ross (1994) has suggested that pre-PMP might have departed from the southeast coast of Taiwan, the Amis language area, since this language name appears to derive from PAn *qamis* 'north'. It is possible that the Amis might have been given this name by the Malayo-Polynesian speakers to the south who might have remembered them as their relatives. Indeed, on linguistic grounds, Reid (1982) considers that an Amis-Extra Formosan node is required in the Austronesian family tree, as follows:
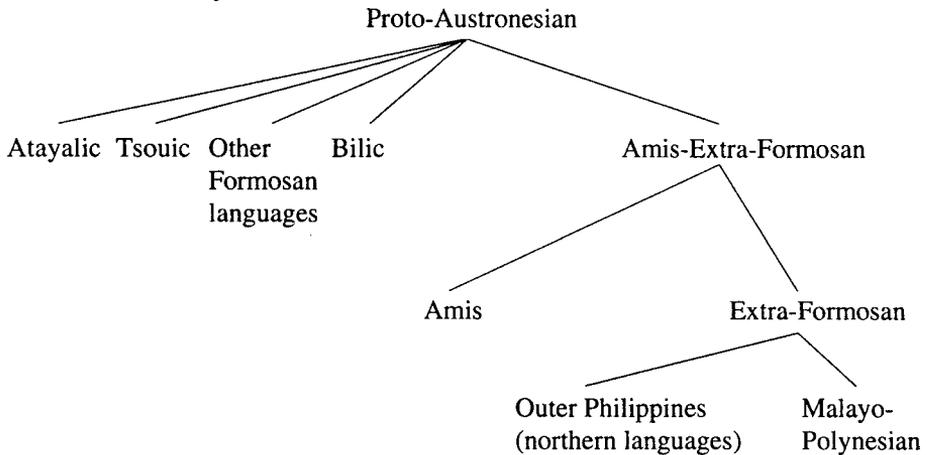
We will return to Reid's assessment of higher level Austronesian subgroups below. First, however, let us return to Blust's Malayo-Polynesian subgroup and its major constituent subgroups, thus:

```
                    Proto-Malayo-Polynesian
                         /         \
                        /           \
              Western              Central-Eastern
          Malayo-Polynesian      Malayo-Polynesian
                                     /       \
                                    /         \
                            Central            Eastern
                       Malayo-Polynesian    Malayo-Polynesian
```

Each of the right hand nodes in the tree diagrams presented here represent the speech of a segment of the population which has migrated from a settled area, so that a new language arose by divergence as a result of separation. Together, the right hand nodes represent the main migratory path of the Austronesians from Taiwan to Oceania. It has been noted, however, that most of the left hand branches do not appear to represent a discrete proto-language, since they represent the "stay-at-homes" (Ross 1994). His further comment on the left-branching nodes is worth quoting in full:

> It looks as if the settled proto-language had already diversified into a local linkage before separation occurred. In these cases, the dialects or languages of the "stay-at-homes" have no exclusively shared ancestor. Instead they share only an ancestor at the node above, with the language of the departed migrants.

The Western Malayo-Polynesian languages include the languages of the Philippines and western Indonesia, including Chamorro, Palauan, Chamic and Malagasy. We know little about the subgrouping of the Western Malayo-Polynesian languages, and as Blust (1985) indicates, there is no clear evidence that these languages form a single subgroup of Austronesian. He is not alone in his thinking.

In fact, there is not even any real agreement as to how the Western Malayo-Polynesian languages subgroup among themselves. Ruhlen (1987), basing himself mainly on the work of Blust, assigns the members of the WMP subgroup to eleven divisions, as follows:
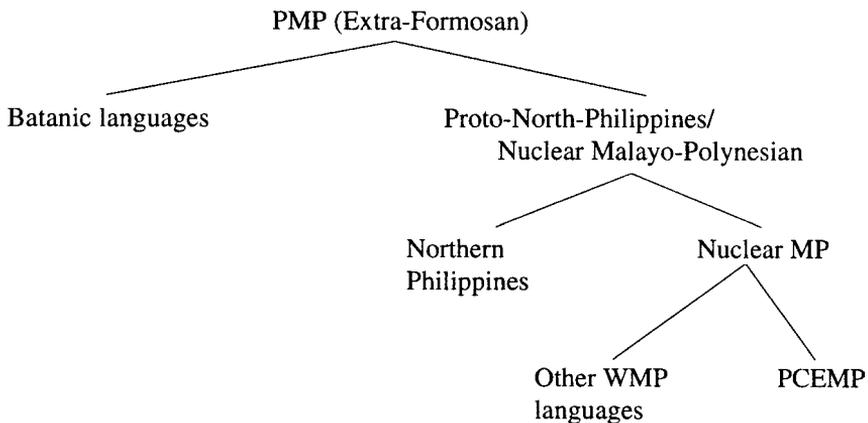
1. Chamorro
2. Palauan
3. Yapese
4. Northern Philippines
5. Southern Philippines
6. Meso-Philippine
7. South Mindanao

8. Sulawesi
9. Borneo
10. Sama-Bajau
11. Sundic

Ruhlen provides no justification for these subgroups, other than a geographical one. It should be noted that Chamorro, Palauan and Yapese are spoken in Micronesia to the east of the Philippines. While Chamorro and Palauan are clearly non-Oceanic, the position of Yapese is less clear.

The languages of the Philippines archipelago (including the Batan Islands between Taiwan and the Philippines) and several groups of languages spoken in the northern arm of Sulawesi have generally been believed to belong to a single Philippines subgroup regarded as having descended from Proto-Philippines (Zorc 1977, 1986). This subgrouping has been assumed rather than justified, however. Reid (1982:202) points out that the innovations which Charles (1974) lists as shared by the languages of the Philippines subgroup are based on a number of phonemic contrasts for PAn proposed by Dyen and Dempwolff which "do not stand close scrutiny, and are probably the result of unrecognized borrowing or obscured phonological processes in the history of the languages involved".

Basing himself on Reid, Ross (1994) suggests another possible scenario, as follows:

PMP (Extra-Formosan)
├─ Batanic languages
└─ Proto-North-Philippines/
   Nuclear Malayo-Polynesian
   ├─ Northern Philippines
   └─ Nuclear MP
      ├─ Other WMP languages
      └─ PCEMP

Reid (1982:212) was also unhappy about the southern Mindanao languages Blaan and Tboli (which continue to exercise his mind today, see below), quite apart from the higher order subgroups linking the languages of Taiwan and the Philippines. He found that the southern Mindanao languages reflected none of the innovations characteristic of the Malayo-Polynesian languages and may be descended from "a very early migration south of Formosa by an Austronesian-speaking people". We will return to this point in a moment, for an update on Reid's current thinking.

Zorc (1986) challenges Reid's subgrouping and defends the notion of a single Philippines grouping. His claim is based on a large number of lexical innovations shared widely by the languages of the Philippines archipelago. Commentators have remarked that it is difficult to assess Zorc's position because it is not clear that his lexical innovations are not in fact vocabulary items that have been retained from PMP but lost in extra-Philippines languages. Reid himself (1982:212) comments: "As one moves south in the Philippines … the degree of influence of one or more of the central Philippines languages becomes more and more pervasive, so that it becomes more and more difficult to separate the strata in the languages".

Reid's thinking today has changed a little, but still centres to some extent around the problem posed by the languages of the Central Philippines, which appear to share a number of innovations with the Malayo-Javanic languages, including the formation of a set of ligatures exclusively shared by these two groups. His current position (pers.comm.) may be represented diagrammatically as follows:
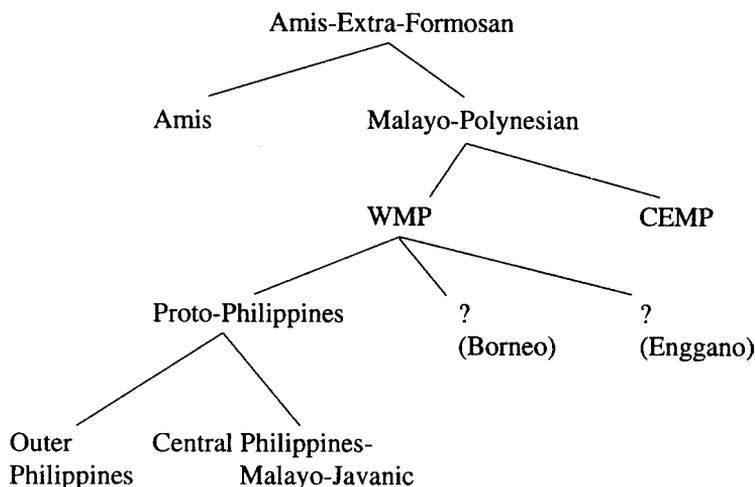
Reid is not so much concerned with the higher level subgroups in this tree diagram as the lower level attempt to resolve the problem posed by the languages of the Central Philippines and their obvious connection with the Malayo-Javanic languages, most probably through a southerly migration or series of migrations.

There are a number of other established subgroups in the Western Malayo-Polynesian area. Blust (pers.comm. to Ross) recognizes the following:

1. Moklen (on islands off the west coast of Thailand and Burma)
2. Lampung (SE Sumatra)
3. Land Dayak (Inland SW Borneo)
4. Southern Philippine/Sangir/Minahasan (Mindanao, N.Sulawesi)
5. Meso Philippine/Mongondow-Gorontalo (C.Philippines, N.Sulawesi)

6. Sama-Bajau
7. NW Sumatra/Barrier Islands (Gayo, Batak, Mentawi, Enggano)
8. NW Borneo
9. Central Sulawesi
10. Southern Sulawesi
11. Tamanic (Central Borneo)
12. Muna-Buton (off SE Sulawesi)
13. Malayo-Chamic (Acehnese, Chamic, Malayan, Sundanese)
14. Java-Bali-Sasak
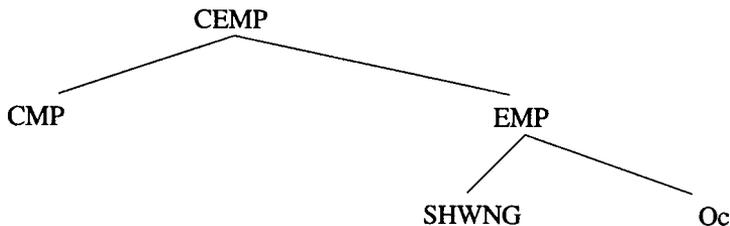15. Barito (Southern Borneo, Madagascar).

There has been further progress made with a number of these proposed lower order subgroups in recent times, as follows:

a. The Malayo-Chamic subgroup, covering substantial areas of western Indonesia (Blust 1985) is evidently closely related to Reid's Central Philippines-Malayo-Javanic subgroup, although details remain to be determined. For work on this subgroup, see also Adelaar (1985), Cowan (1948, 1974) and Durie (1989). For Malayo-Javanic see Nothofer (1986, 1991, 1994). In this context the wider relationships of the languages of the Java-Bali-Sasak subgroup proposed by Esser (1938) remain to be worked out. While the Bali-Sasak connection is well established, however, the relationship of these two languages with Javanese has not yet been formally demonstrated.

b. The Tamanic subgroup has been established by Adelaar (1994 and this volume), linking the Tamanic languages of central Borneo and the languages of southern Sulawesi. The languages of southern Sulawesi had previously been shown to form a subgroup by Mills (1975).

c. The Barito subgroup, on South Borneo, is best known because of one of its members, Malagasy. Until recently Malagasy was believed to have been in Madagascar since the fourth-fifth century. Adelaar (1994) suggests that this date is at least two centuries too early.

Nothofer (1990, 1994) has made a number of fresh proposals concerning Western Malayo-Polynesian. His proposal is that much of the WMP region was once occupied by speakers of languages belonging to a group which he calls "Palaeo-Hesperonesian", and that at a later date much of this area was occupied by speakers of "Hesperonesian" languages, who became culturally dominant in western Indonesia, displacing the Palaeo-Hesperonesian languages. Some of these survive today around the periphery of the WMP region. In Nothofer's terms, the languages of northern Sulawesi and the central and southern Philippines, together with those of north-west Sumatra and the Barrier Islands, north-west Borneo and central and southern Sulawesi would be Palaeo-Hesperonesian, while the Malayo-Chamic, Java-Bali-Sasak and Barito

groups are Hesperonesian. Ross (1994) notes, however, that much of the evidence which Nothofer uses is lexical and so suffers from the same difficulties as Zorc's use of such evidence in the Philippines. By the same token, however, Nothofer's hypothesis must be considered seriously.

The Central-Eastern Malayo-Polynesian subgroup of Austronesian is much more substantial. It was first proposed as Eastern Austronesian by Blust (1974), and later rebaptized Central-Eastern Malayo-Polynesian (CEMP) by the same scholar (Blust 1978). The languages which constitute the CEMP subgroup stretch from Bimanese, on the island of Sumbawa, eastward through the Lesser Sunda chain of Indonesia as far as the Aru Islands, and then north-west into the central Moluccas, inclusive of the Sula Archipelago. In addition, several still very poorly known CEMP languages appear to be scattered along the southern coast of Irian. CEMP and its lower order subgroups are as follows:

```
                        CEMP
                       /      \
                      /        \
                  CMP           EMP
                               /    \
                              /      \
                         SHWNG        Oc
```

Blust (1990:2) states that "the evidence for CEMP and for some previously unrecognized subgroups within CMP is considerably stronger than the evidence for CMP itself". For, as we shall see, CMP is "united" by a number of overlapping innovations which cover many, but not all of the languages in question. This distribution of non-coincident innovations suggests to Blust that at an early stage in the Austronesian settlement of eastern Indonesia the languages now assigned to CMP formed a relatively isolated dialect chain which still shared well over 90 per cent of its basic vocabulary with languages that were not part of that chain.

In terms of culture-historical implications, after its separation from Proto-Western Malayo-Polynesian (PWMP), Proto-Central-Eastern Malayo-Polynesian (PCEMP) developed for some time in a relatively compact geographical area before splitting into Proto-Central Malayo-Polynesian (PCMP) and Proto-Eastern Malayo-Polynesian (PEMP). PEMP and its immediate descendant, Proto-Oceanic (POc) each developed in a relatively compact geographical area before splitting into descendant languages. By contrast, Proto-South Halmahera-West New Guinea (PSHWNG) and PCMP spread rapidly over a considerable distance before much dialect differentiation existed. A large number of linguistic innovations arose and spread through the CMP dialect chain in opposite directions, as they did also in SHWNG. These changes failed to reach the geographical extremes furthest from their respective centres of origin, producing differences of rule ordering in the central diffusion corridor (Blust

1978). The result is a patchy distribution of widely dispersed innovations. On the other hand, Blust maintains that some recurrent changes in the CMP languages may have been independent of contact, hence products of drift. Finally, after the differentiation of the CMP chain into distinct languages, there were limited migrations of small populations from the southern Moluccas in Indonesia to the southern coast of the Bird's Head Peninsula of New Guinea.

The evidence for the existence of a CEMP subgroup by Blust is quite substantial. It consists of the following:

1. The reduction of consonant clusters in reflexes of reduplicated monosyllables.
2. Irregular phonological change in five lexical items, 33 apparent lexical innovations, two broad structural agreements that appear to involve innovations, irregular morphological changes in four lexical items and seven semantic innovations.

Blust maintains that there is little to distinguish PMP from PCEMP phonologically. PMP *c and *s apparently merged as *s. But a similar merger occurs in many WMP languages and in all Formosan languages. However, as mentioned above, there is a reduction of hetero-organic consonant clusters in the reflexes of reduplicated monosyllables. All CEMP languages have simplified medial clusters in the reflexes of PMP reduplicated monosyllables (unless the cluster consisted of a nasal followed by a stop or fricative, in which case the nasal assimilated to the place of articulation of the stop, but was not lost). Examples:

PMP *bukbuk > PCEMP *bubuk	'wood weevil'
PMP *ñamñam > PCEMP *ñañam	'tasty, delicious'
PMP *mekmek > PCEMP *memek	'crumbs'

Some WMP languages, for example Malay, have made similar simplifications — but, in Blust's opinion, the universality of this change in CEMP is best explained as the product of a single innovation in a language ancestral to the whole group.

Further evidence for the CEMP subgroup cited by Blust (1990) is as follows:

1. Irregular phonological change in five lexical items:

PMP *uliq	PCEMP *oliq	'return, go back'
PMP *i-sai	PCEMP *i-sei	'who?'
PMP *ma-qitem	PCEMP *ma-qet əm	'black'
PMP *maRi	PCEMP *mai	'come'
PMP *tudan	PCEMP *todan	'sit'

2. Apparent lexical innovations in PCEMP

1.	PMP *ka-labaw	PCEMP *kanzupay	'rat'
2.		PCEMP *liqə	'voice'
3.		PCEMP *malu	'loincloth'
4.	PMP *dilaq	PCEMP *maya	'tongue'
5.	PMP *surat	PCEMP *tusi	'scratch, draw a line, etc.'
6.	PMP *tawa	PCEMP *malip	'laugh'
7.		PCEMP *saRa	'sweep, broom'
8.		PCEMP *kandoRa	'cuscus, phalanger'
9.		PCEMP *mansar/mansər	'bandicoot'
10.		PCEMP *keRa(ŋ)	'hawksbill turtle'
11.	PMP *amuR	PCEMP *au	'dew'
12.		PCEMP *bai	'do, make'
13.	PMP *paen	PCEMP *bayan/payan	'bait'
14.	PMP *hazani	PCEMP *da ŋi	'near'
15.		PCEMP *kese	'keep to oneself, be different'
16.	PMP *dalem	PCEMP *laman	'deep'
17.	PMP *paen	PCEMP *pani(n ŋ)	'bait'
18.	PMP *muRmuR	PCEMP *pupuR	'gargle, rinse the mouth'
19.	PMP *kapal	PCEMP *t əlu	'thick (of materials)'
20.		PCEMP *qumun	'earth oven'
21.	PMP *lakaw/panaw	PCEMP *ba	'go'
22.		PCEMP *balaŋ	'side, part'
23.	PMP *qa-lima	PCEMP *baRa	'hand, arm'
24.		PCEMP *lama	'spread over, cover'
25.		PCEMP *ŋaRa	'wild duck'
26.		PCEMP *papaR	'cheek'
27.		PCEMP *paRa-	'reciprocal prefix'
28.	PMP *palihi	PCEMP *tambu	'forbid'
29.	PMP *hiup	PCEMP *upi	'to blow'
30.		PCEMP *waŋka	'canoe'
31.		PCEMP *wari	'sing, song'
32.	PMP *ma-esak	PCEMP *madar	'ripe, overripe'
33.	PMP *bahu	PCEMP *mapu	'unpleasant odour'

3. Structural agreements:

Blust (1990) maintains that there are two features which are widely distributed in Eastern Indonesia and Oceania, namely:

a. The use of proclitic subject markers on the verb.
b. A morphologically marked distinction between alienable and inalienable possession.

However, there is a lack of established cognation in the morphemes used to express formally similar systems — thus a hypothesis of convergent development between the CMP and the Oc proclitics cannot easily be ruled out. Indeed Ross (1988:96ff.) also questions whether there is convincing evidence for an immediate common ancestor of the CMP, SHWNG and Oc subgroups.

4. Irregular morphological changes:

1.	PMP *apa	PCEMP *sapa	'what?'
2.	PMP *hepat	PCEMP *pat, *pati	'four'
3.	PMP *ma-huab	PCEMP *mawab	'yawn'
4.	PMP *ma-hiaq	PCEMP *mayaq	'shy, ashamed'

Blust (1990) concludes that the evidence for the existence of the PCEMP subgroup is fairly strong, as individual pieces of evidence are mostly mutually independent. Grimes (1990) has made an independent evaluation of the CEMP evidence and finds that Blust has a good case, even though very few of the lexical innovations which Blust lists are replacement innovations.

With respect to the Central Malayo-Polynesian subgroup (CMP), Blust and others are much less confident. These are the languages of the Lesser Sundas east of the Bima-Sumba group, and those of the southern and central Moluccas. The problems associated with this subgroup are not surprising, as we are again dealing with a "stay-at-home" rather than a migratory group. As suggested above, the most striking feature of the phonological history of the CMP languages is the extent to which similar changes are found in many but not all of the languages. This pattern of innovation suggests that PCMP underwent a short period of development apart from other contemporary Austronesian languages before it began to spread from the Moluccas to the Lesser Sundas. Many of the changes that are now widespread in these languages took place after this geographical dispersal and were the result of diffusion and in some cases drift.

The innovations which distinguish the CMP languages according to Blust (1990) are the following:

1. Loss of the prepenultimate initial vowel *hV- or *qV-. PMP trisyllables which began with a vowel or a vowel preceded by a laryngeal *h- or *q- (e.g. *qateluR 'egg') were retained as such in POc, but cognate forms in

CMP languages suggest that the first syllable was lost in PCMP. However, Blust admits that forms such as Watubela /katlu/ make this claim rather doubtful.

2. Glide truncation: The monophthongization of original diphthongs through the truncation of glides, as in *-ay > -a; *-aw > -a; *-uy > -u is one of the most distinctive characteristics of the phonological history of CMP languages. This change is unattested outside eastern Indonesia. However, Blust concedes that we are forced to conclude that this was a product of independent changes in a number of languages. Truncation appears only sporadically in Leti, Kisar and Erai and is apparently unknown in Timor, Roti, Savu or Sumba.

3. Postnasal voicing: In many of the CMP languages, stops have become voiced after nasals. This is true of both consonant clusters within a morpheme and of intermorphemic clusters created by syncope. Thus, for example:

PMP *ma-putiq > Kemak (C.Timor), Bonfia (E.Seram) buti, Buru boti, 'white'.

As with the two previously discussed innovations, however, it appears that postnasal voicing was also a recurrent change.

4. Irregular sound changes characterize a number of the CMP languages:
   a. PMP *pandan PCMP *pendan 'pandanus'

      However, while examples of this irregular development are known from Flores to the Leti-Moa Archipelago, they are apparently not found in the southern and central Moluccas.
   b. PMP *baqeRu PCMP *beqeRu 'new'

5. Lexical Innovations:

   Blust lists the following innovations which he claims are exclusively shared by the languages of the Lesser Sundas and the Moluccas:

1.		PCMP *balabu	'see dimly'
2.		PCMP *balik	'mix, blend'
3.		PCMP *beta	'cut wood'
4.		PCMP *dada	'drag'
5.		PCMP *dodok	'pierce, stab'
6.		PCMP *letay	'above'
7.	PMP *kawit	PCMP *gae	'hook'
8.		PCMP *kati	'call a dog'
9.		PCMP *ketu	'pluck, break off'
10.		PCMP *lemba	'carry with a carrying pole'
11.		PCMP *lesi	'excess, overabundance'
12.		PCMP *lesu	'come out, take out'
13.		PCMP *letay	'bridge'

14.		PCMP *leu*	'bend'
15.		PCMP *liRi*	'sound, voice'
16.		PCMP *lolan*	'cut off a piece'
17.		PCMP *lunu*	'roll up'
18.	PMP *i-nu*	PCMP *mpae*	'where?'
19.		PCMP *peu*	'bind together in a sheaf'
20.	PMP *qasu*	PCMP *masu*	'smoke'
21.		PCMP *silu*	'lift, raise'
22.	PMP *tahiq*, *zaqit*	PCMP *sora*	'sew'
23.		PCMP *sula*	'horn'
24.		PCMP *ta*	'no, not'
25.	PMP *taliŋ a*	PCMP *tilu*	'ear'

The major problem with the lexical innovations as proposed here is that again they are not replacement innovations.

Blust (1990) also proposes some morphosyntactic and semantic innovations for the CMP subgroup, but here again the problem is that they are not shared throughout the proposed subgroup. In fact that is the very point which Blust himself makes. Blust asks whether we should assume that the changes he has documented are the product of completely independent innovations, that is, of drift. If so, he says, it is puzzling why the changes in question should be concentrated in the languages of the Lesser Sundas and the southern and central Moluccas.

In Blust's opinion, diffusion is the most plausible explanation for the distributions he puts forward. It is well known that diffusion can occur across major subgroup boundaries. Thus the widely shared phonological innovations among the languages of the Lesser Sundas, the southern and central Moluccas and the southern part of the Vogelkop Peninsula may simply be the products of contact among Austronesian languages that share no particularly close genetic affinity.

The Central Malayo-Polynesian subgroup of Austronesian, then, is faced with the same kinds of problems as other ''stay-at-home'' Austronesian groups, and its existence cannot at this stage be taken as proven any more than that of the Western Malayo-Polynesian subgroup. Nobody has really looked at the over-all relationships of the languages of Nusatenggara and Timor with the languages of Maluku. Thus we have no real idea of the first-order nodes within CMP.

The two descendants of the Eastern Malayo-Polynesian subgroup are the South Halmahera-West New Guinea (SHWNG) and Oceanic (Oc) subgroups. The

SHWNG subgroup consists of all of the Austronesian languages of Halmahera and its near satellites and the various languages along the north coast of the Vogelkop Peninsula and Cenderawasih Bay, Waropen and all the Austronesian languages of Yapen Island and its satellites. The data available for many of these languages is far from adequate, making subgrouping difficult and at present uncertain. One important problem remaining to be solved concerns the boundary between CMP and SHWNG languages.

Blust (1978) set out the criteria for the SHWNG and Oc subgroups of Austronesian and need not be repeated in full here. In summary, Blust considers that the following are the most useful defining innovations for the South Halmahera-West New Guinea subgroup:

1. Postnasal syncope (loss of a vowel between a nasal and a following stop, e.g. PMP *mata* > PSHWNG *mta* 'eye').
2. The shift of PMP *e* > PSHWNG *o* in penultimate syllables.
3. The replacement of PMP *anak* > PEMP *natu* 'child'.

In terms of the Oceanic subgroup (Oc), Ross (1988:30) sets out a list of ten phonological innovations which distinguish POc (Proto-Oceanic) from PAn (Proto-Austronesian). However, half of these are also reflected in the South Halmahera-West New Guinea subgroup (SHWNG) and as such are attributable to Proto-Eastern Malayo-Polynesian (PEMP), the immediate ancestor of both SHWNG and POc. There are, however, five innovations shared exclusively between PEMP and POc, as follows:

	PEMP		POc
1.	*(m)p*	*(m)b*	(m)p, ŋp
2.	*(n)s*	*(n)z*	(n)s
3.	*e*	*aw*	o
4.	*ay*	*ey*	e
5.	*m*		m, ŋm

In terms of phonological innovations between PEMP and POc, then, we are dealing with four mergers and two splits, quite substantial evidence by any standards. There is also lexical and morpho-syntactic evidence for the existence of the Oceanic subgroup presented in Pawley (1972:2-3). The development and dispersal of the Oceanic subgroup of Austronesian is discussed in the following chapter by Pawley and Ross.

Cultural-Historical Implications

The first Austronesians are believed to have originated in the South China area before moving off from the Asian mainland to settle on Taiwan somewhere about 5000-6000 years ago. There they lived relatively undisturbed for some time

before one of the Taiwan-Austronesian communities, possibly from the south-east, moved south to the Philippines. Eventually they moved right through the Philippine archipelago. From there one group moved south-west, through Borneo and later Sumatra and Java, with branches penetrating the Malay Peninsula, eastern parts of Vietnam and Cambodia. A second migration from the Philippines moved south into Sulawesi. From there it is believed to have followed two major paths, one through Sulawesi and into the Seram-Ambon area and Timor, and the other towards Halmahera and Irian Jaya. From there the Austronesians are believed to have moved eastwards along the north coast of Papua New Guinea, ending up in the Bismarck Archipelago (New Britain and New Ireland), where the pre-Proto-Oceanic community is considered to have remained relatively unmolested until they were ready to move out into the Pacific.

References

Adelaar, K.A.

1981   Reconstruction of Proto-Batak phonology. *NUSA: Linguistic Studies in Indonesian and Languages in Indonesia* 10:1-20.

1985 Proto-Malayic: the reconstruction of its phonology and parts of its lexicon and morphology. PhD dissertation, University of Leiden.

1994   The classification of the Tamanic languages. In Tom Dutton and Darrell Tryon (eds) *Contact-induced change in Austronesian languages*, pp.1-41. Berlin: Mouton de Gruyter.

Blust, Robert A.

1974   Eastern Austronesian: a note. *Working Papers in Linguistics, University of Hawaii* 6(4):101-107.

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

1978   Eastern Malayo-Polynesian: a subgrouping argument. In S.A. Wurm and Lois Carrington (eds) *Second international conference on Austronesian linguistics proceedings*, Fascicle 1, Pacific Linguistics Series C No. 61, pp.181-234. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

1982   The linguistic value of the Wallace line. *Bijdragen tot de Taal-, Lande- en Volkenkunde* 138(2-3):231-250.

1983-84   More on the position of the languages of eastern Indonesia. *Oceanic Linguistics* 22-23:1-28.

1985   The Austronesian homeland: a linguistic perspective. *Asian Perspectives* 26(1):45-67.

1990   Central and Central-Eastern Malayo-Polynesian. Paper presented at Conference on Maluku linguistics, University of Hawaii, March 1990.

Charles, M.

1974   Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics* 13:457-509.

Cowan, H.J.K.

1948   Aanteekeningen betreffende de verhouding van het Atjehsch tot de Mon-Khmer-talen. *Bijdragen tot de Taal-, Lande- en Volkenkunde* 104:429-514.

1974   Evidence of long vowels in early Acehnese. *Oceanic Linguistics* 13:187-212.

Dempwolff, Otto

1934-38   *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols. Berlin: Reimer.

Durie, Mark

1989   Proto-Chamic and Acehnese mid vowels: towards Proto-Aceh-Chamic. *Bulletin of the School of Oriental and African Studies* 53:100-114.

Dyen, Isidore

1965   *A lexicostatistical classification of the Austronesian languages*. Bloomington: Indiana University Publications in Anthropology and Linguistics, Memoir 19; supplement to *International Journal of American Linguistics* 25.

1990   *Homomeric lexical classification*. In Philip Baldi (ed.) *Linguistic change and reconstruction methodology (Trends in Linguistics. Studies and Monographs 45)*, pp. 211-229. Berlin/New York: Mouton de Gruyter.

Esser, S.J.

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

Grimes, C.E.

1990   Notes on Blust 1990. (Mimeo.)

Li, Paul Jen-kuei

1980   The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology* 51:349-405. Taipei: Academia Sinica.

1981   Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology* 52:235-301. Taipei: Academia Sinica.

1985   The position of Atayal in the Austronesian family. In Andrew Pawley and Lois Carrington (eds) *Austronesian linguistics at the 15th Pacific science congress*, pp.257-280. Pacific Linguistics Series C No. 88. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Mills, Roger F.

1975   Proto-South Sulawesi and Proto-Austronesian phonology. PhD dissertation, University of Michigan. Ann Arbor: University Microfilms International.

Nothofer, B.

1986   The Barrier Island languages in the Austronesian language family. In Paul Geraghty, Lois Carrington and S.A. Wurm (eds) *FOCAL II: Papers from the fourth international conference on Austronesian linguistics*. Pacific Linguistics Series C No. 94, pp.87-109. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

1991   Current interpretations of Western Malayo-Polynesian linguistic prehistory. *Bulletin of the Indo-Pacific Prehistory Association* 12:388-397.

1994   The relationship between the languages of the Barrier Islands and the Sulawesi-Philippine languages. In Tom Dutton and Darrell Tryon (eds) *Contact-induced change in Austronesian languages*, pp.389-409. Berlin: Mouton de Gruyter.

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*, Vol. 3. Honolulu: Pacific Anthropological Records No. 13, Bernice P. Bishop Museum.

Reid, Lawrence A.

1982   *The demise of Proto-Philippines. Papers from the third international conference on Austronesian linguistics*, Vol. 2, pp.201-216. Pacific Linguistics Series C No. 75. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Ross, Malcolm D.

1988   *Proto Oceanic and the Austronesian languages of western Melanesia*. Pacific Linguistics Series C No. 98. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

1994   Some current issues in Austronesian linguistics. In Darrell T. Tryon (ed.) *Comparative Austronesian dictionary*, pp.45-120. Berlin: Mouton de Gruyter.

Ruhlen, Merritt

1987 *A guide to the world's languages*. Stanford, California: Stanford University Press.

Stresemann, E.

1927 Die Lautentsprechungen in den ambonischen Sprachen. *Zeitschrift fur Eingeborenen-Sprachen*, Supplement 10, Berlin.

Tryon, Darrell T. (ed.)

1994 *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter.

Tsuchida, S.

1976 *Reconstruction of Proto-Tsouic phonology. Study of languages and cultures of Asia and Africa*. Monograph Series 5, Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Zorc, R. David

1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Pacific Linguistics Series C No. 44. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

1986 The genetic relationships of Philippine languages. In Paul Geraghty, Lois Carrington and S.A. Wurm (eds) *FOCAL II: Papers from the fourth international conference on Austronesian linguistics*, pp.147-173. Pacific Linguistics Series C No. 94. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.