

The Batanic Languages in Relation to the Early History of the Malayo-Polynesian Subgroup of Austronesian*

Malcolm Ross**

ABSTRACT

The Batanic languages are a group of closely related Austronesian languages spoken on the small islands scattered between Taiwan and Luzon. The purpose of this paper is to present what is known of their relationships to other Austronesian groups, the Formosan languages to the north and the Malayo-Polynesian languages to the south. It is shown that the Batanic languages are Malayo-Polynesian, but do not subgroup with any other smaller group of languages within the huge Malayo-Polynesian language group. In view of this, it is possible that the Batanic languages are descended from the speech of a community established in the islands by speakers of Proto Malayo-Polynesian before they reached Luzon. This possibility is supported by the conservatism of the Batanic languages, but remains just that—a possibility, not a certainty.

The paper concludes with a consideration of the question, what is the closest relative within Taiwan of Proto Malayo-Polynesian? and concludes that there is at present no clear answer to this question.

Key Words: Malayo-Polynesian, Batanic languages

1 INTRODUCTION

The Batanic languages⁽¹⁾ are a group of closely related languages spoken on the small islands scattered between Taiwan and Luzon. According to Tsuchida, Yamada and

* My thanks go to Lawrence Reid, whose comments helped immensely in this revision of a paper presented to the Conference *The Asian Fore Arc Project: Results and prospects from the Philippines and Taiwan* (at The Australian National University, 5–6 August 2004). Remaining errors are of course my responsibility.

** Research School of Pacific and Asian Studies and Centre for Research on Language Change, The Australian National University

(1) Sometimes called 'Bashiic' or 'Ivatanic' or 'Vasayic'.

Moriguchi (1987) and Tsuchida, Constantino, Yamada and Moriguchi (1989), there are four languages in the group, listed in Table 1 together with their locations and the dialects for which Tsuchida and his colleagues provide data. The map in Figure 1 shows approximate locations.

TABLE 1: The Batanic languages

Language	Dialects	Location
Yami	Imurud, Iraralay, Iranumilek	Lányü (Orchid Island) (Taiwan)
Itbayat	...	Itbayat (Philippines)
Ivatan	Ivasay, Isamurung	Batan, Sabtang (Philippines)
Babuyan	...	Babuyan (Philippines)

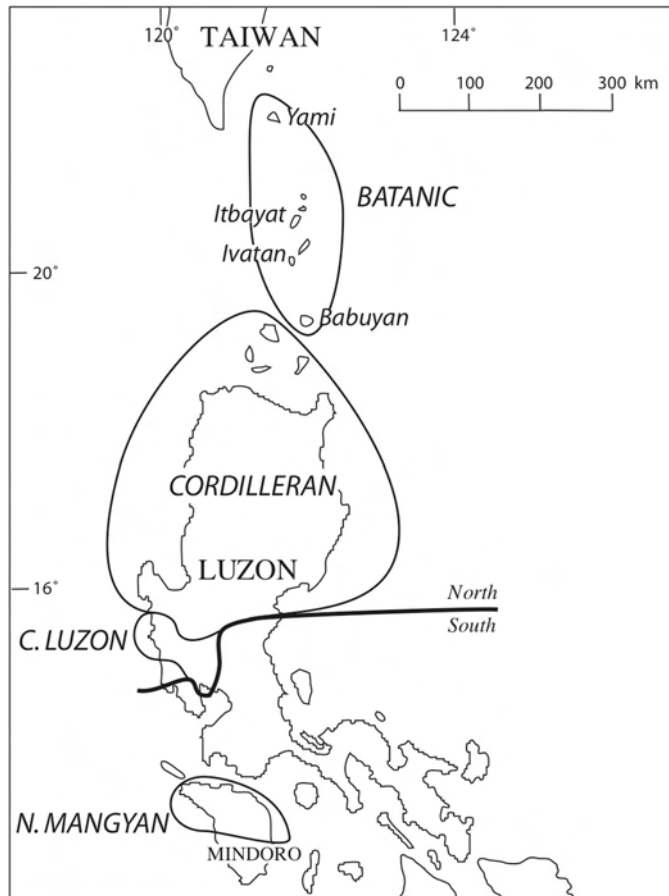


Fig. 1: The Batanic languages

These are Austronesian languages, and the purpose of this paper is to present what is known of their relationships to the Austronesian languages to the north and south. We must recognise that human lineages and linguistic lineages do not always match each other, as human communities may switch from one language to another over time. Some Babuyan speakers are bilingual in Ilokano (northern Luzon), and some of these speak Ilokano as their first language, and this has recently affected the language spoken on Babuyan.⁽²⁾ However, this is the only evidence we have of language shift and language contact in the history of today's Batanic-speaking communities, and this paper is directly concerned only with the history of the Batanic linguistic lineage. Conclusions based on a language about the history of its speakers can be drawn only with caution.

The Austronesian languages to the north of the Batanic group are the Formosan languages, *i.e.* the aboriginal languages of Taiwan. The Austronesian languages to the south are the Cordilleran languages of northern Luzon. In order to position the Batanic languages in the Austronesian family, it is necessary first to understand where the Formosan and Cordilleran groups fit into the wider relationships of the family, and this is the subject of §2.

One thing can immediately be said about the relationships of the Batanic languages to one another. On the basis of the available data they appear so similar to each other that one can readily reconstruct Proto Batanic, the language from which they are all immediately descended. No other known languages are daughters of Proto Batanic.

Indeed, it is questionable whether Itbayat, Ivatan and Babuyan should be regarded as separate languages. Scheerer (1926:301), Hidalgo (1996:22) and Reid (1966: Preface) all treat Itbayat as a dialect of Ivatan, although Scheerer comments that Ivatan speakers cannot understand Itbayat. Hidalgo claims a lexicostatistical relationship of 87% between Ivatan and Itbayat, easily a dialect relationship.⁽³⁾ Reid distinguishes three Ivatan dialects, northern, central and southern. His northern dialect is Itbayat, the central dialect is spoken in the town of Basco on Batan Island, and the southern dialect is spoken throughout the rest of Batan and on Sabtang Island. The central and southern dialects apparently correspond to the Ivasay and Isamurung dialects listed in Table 1.

There is historical and ethnographic evidence to suggest that Ivatan and Babuyan were until recently a single speech community. Spanish records indicate that the earlier population of Babuyan was removed by the colonial power, and that it was only at the end of Spanish rule that Babuyan was repopulated by a few families, most of them speakers of one of the Ivatan dialects of Batan Island.⁽⁴⁾

(2) Notes by Rundell Maree (2003) of the Philippine Branch of the Summer Institute of Linguistics, made available to me by Lawrence Reid.

(3) However, Hidalgo does not indicate the wordlist that this figure is based on.

(4) Notes by Rundell Maree (see footnote 3). See also Maree (1982).

The Batanic data used in the small study reported here are drawn mainly from Tsuchida *et al.* (1987) and Tsuchida *et al.* (1989). These data compilations were collected in order to determine the position of the Batanic languages in relation to other Austronesian languages, but the authors do not draw any conclusions themselves. This paper attempts to fill that gap.

Tsuchida *et al.* (1987) is a comparative list of over eight hundred words for the four languages. Data were collected for the Imorod and Iraralay dialects of Yami, for Itbayat, for the Ivasay and Isamurung dialects of Ivatan, and for Babuyan.

Tsuchida *et al.* (1989) is a comparative list of nearly six hundred phrases and sentences for the four languages. Data were collected from the same dialects as for Tsuchida *et al.* (1987), except that the Iraralay dialect of Yami is replaced by the Iranumilek dialect.

Data on verbal morphology are drawn from Shih (1997) for Yami (Imurud dialect) and from Reid (1966) for Ivatan (central, apparently Ivasay, dialect).

2 AUSTRONESIAN RELATIONSHIPS AND PROTO BATANIC

The upper nodes of the Austronesian family tree, in a form probably accepted by a majority of Austronesian historical linguists and based on the work of Blust (1977, 1999), are shown in Figure 2. The groups to which the tree refers are shown in the map in Figure 3.

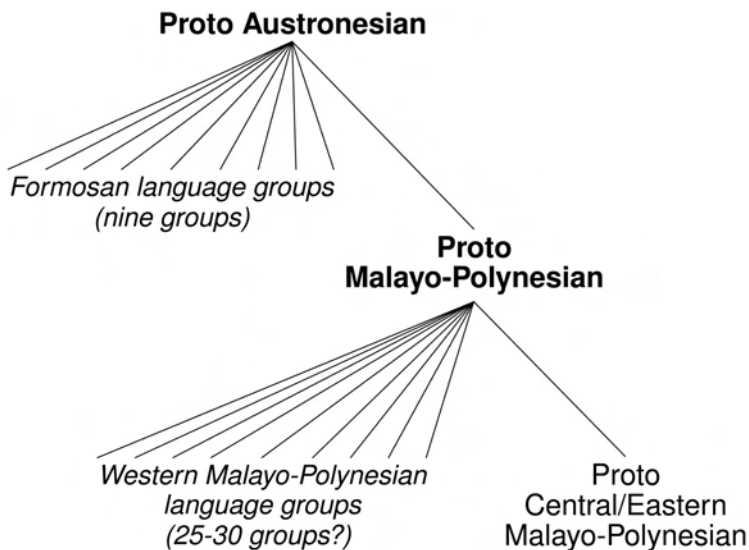


Fig. 2: The upper nodes of the Austronesian family tree

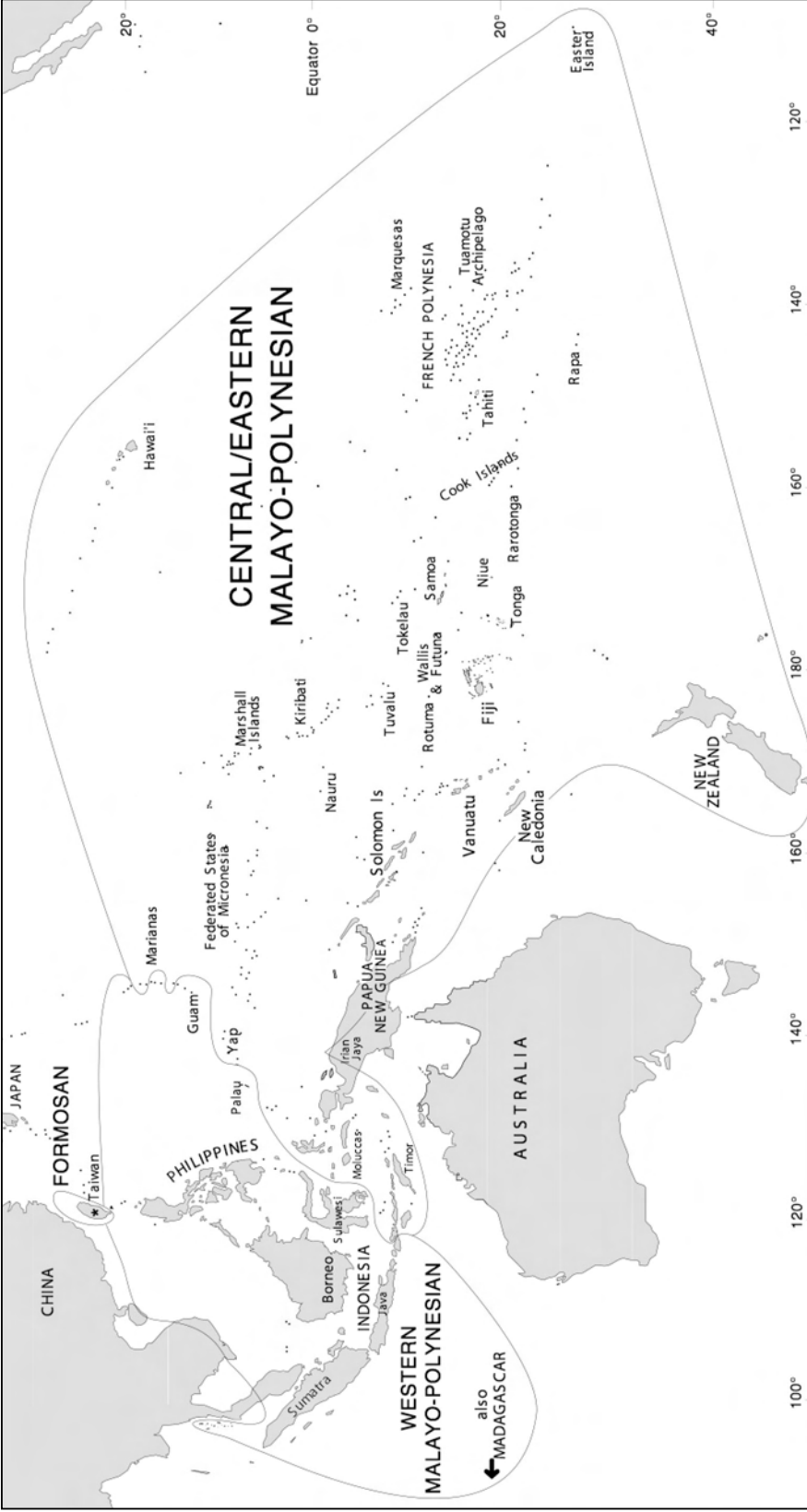


Fig. 3: Major Austronesian groupings: see Fig. 2

The procedure used to determine a family tree consists first of comparing the languages of the family and reconstructing the protolanguage from which they are all descended, *i.e.* in this case Proto Austronesian, and secondly of identifying (sub)groups of languages whose members share a set of innovations relative to the reconstructed protolanguage. If a set of innovations is shared by the languages of a group, it is inferred that they are shared because they have been inherited from a single interstage language. This is far more probable than the alternative assumption—that the innovations have occurred independently in each language which reflects them.⁽⁵⁾

Thus in Figure 2, Proto Austronesian is shown as having ten daughter groups. One of these is the Malayo-Polynesian group, which includes **all** Austronesian languages spoken outside Taiwan. The other nine are groups of Formosan languages spoken (or formerly spoken) in Taiwan.⁽⁶⁾ The interstage language from which all Malayo-Polynesian languages are descended is shown in Figure 2 as Proto Malayo-Polynesian. The nine putative interstage languages from which the known Formosan languages are descended are not shown.⁽⁷⁾

The shape of the tree in Figure 2 suggests a pattern of dispersal which occurred twice.⁽⁸⁾ First, speakers of Proto Austronesian spread across Taiwan. Initially dialect variation arose in the protolanguage. Over time, this variation led to loss of mutual intelligibility between dialects, *i.e.* under a definition often used by linguists the dialects became separate languages. Some of these were the nine interstage languages just mentioned. Note that each Formosan group is directly descended from Proto Austronesian: there was no single 'Proto Formosan' language. Around 4000 BP on current archaeological dating, one group of speakers migrated out of Taiwan in a southerly direction, perhaps to one or more of the islands where the Batanic languages are now spoken, perhaps to Luzon. Their speech underwent certain innovations and became Proto Malayo-Polynesian.

This pattern was repeated during the dispersal of Proto Malayo-Polynesian. Speakers spread, perhaps quite rapidly, occupying suitable ecological niches in the Philippines, Borneo, Sulawesi, the Malay Peninsula, Sumatra and Java, then variation again led to loss of mutual intelligibility and many separate languages, some of which were the interstage languages from which the present-day 20–25 groups of western Malayo-Polynesian languages are descended.⁽⁹⁾ The literature often refers to a discrete 'Proto Western Malayo-Polynesian',

(5) This is a highly simplified account of the linguist's comparative method, but it is sufficient for the present purpose.

(6) When westerners and the Southern Min arrived in the seventeenth century, there were presumably more Formosan languages of which no record has been preserved.

(7) Recent work by Laurent Sagart 2004 suggests that Tai-Kadai belongs as an eleventh group in this tree, but this is not relevant to the theme of this paper.

(8) For a more detailed account, see Pawley (2002).

(9) A more detailed list of possible groups is given by Ross (1995), but this is now somewhat out of date.

but there is no evidence (in the form of shared innovations) that such a language ever existed.⁽¹⁰⁾ Sometime after 4000 BP but before 3500 BP (again, on archaeological dating) a community of speakers somewhere on the southeastern periphery of the Malayo-Polynesian expansion became separated from fellow speakers, and their speech evolved into Proto Central/Eastern Malayo-Polynesian, the language ancestral to the rest of the Austronesian language family—the languages of the eastern Indonesian region, New Guinea, Island Melanesia, Polynesia and Micronesia. (The genesis of Proto Central/Eastern Malayo-Polynesian is ill-understood and may have been more complicated than this, but this does not affect the history of the Batanic languages.)

Where does Proto Batanic fit into the tree in Figure 2? The answer is that it reflects the innovations which Malayo-Polynesian languages share, and is one of the 20–25 western Malayo-Polynesian groups shown in Figure 2.

Innovations which occurred in Proto Malayo-Polynesian (PMP) relative to Proto Austronesian (PAn) are both phonological and morphological. I enumerate them here for the sake of explicitness, but keep discussion and examples to a minimum.

The phonological innovations are (Blust 1990):⁽¹¹⁾

(1) a. PAn *t and *C merged as PMP *t.⁽¹²⁾

b. PAn *L and *n merged (with some unexplained exceptions) as PMP *n.⁽¹³⁾

c. PAn *S and *h merged as PMP *h.⁽¹⁴⁾

Innovation (1a) is illustrated in (2), where both PAn *t and PAn *C are reflected as Itbayat *t* (the forms in other Batanic languages are almost identical):

(2) a. PAn *tuLa ‘freshwater eel’ >PMP *tuna >Itbayat *tuna*

b. PAn *pitu ‘seven’ >PMP *pitu >Itbayat *pitu*

(10) The languages of different western Malayo-Polynesian groups often look quite similar to each other, but there is good evidence that this is the outcome of shared retentions, not shared innovations.

(11) I have argued (Ross 1992) for for two further PMP phonological innovations, namely the merger of PAn *d₁, *d₂ and *d₃ as PMP *d and the acquisition by PMP of a palatal nasal *ñ. Blust (1999) argues against these. With regard to the first, I now agree with him. With regard to the second, I remain reluctant to follow him in reconstructing PAn *ñ, but I ignore the proposed innovation here.

(12) PAn *t was probably a dental stop [t̪], *C an alveolar affricate [ts] (Ross 1992).

(13) PAn *L was probably dental/alveolar [l], contrasting with retroflex *l̪ [l̪] (Ross 1992).

(14) PAn *S was probably retroflex [ʂ] or alveopalatal [ʃ], contrasting with alveolar *s [s] (Ross 1992).

c. PAn *Caliŋa ‘ear’ >PMP *taliŋa >Itbayat *taliña*

d. PAn *maCa ‘eye’ >PMP *mata >Itbayat *mata*

Innovation (1b) is illustrated in (4), where both PAn *L and PAn *n are reflected as Itbayat *n*:

(3) a. PAn *qaLup ‘hunt’ >PMP *qanup >Itbayat *anup*

b. PAn *dapaL ‘sole’ >PMP *dapan >Itbayat *rapan*

c. PAn *wanan ‘right (hand)’ >PMP *wanan >Itbayat *wanan*

Innovation (1c) is illustrated in (3):

(4) a. PAn *duSa ‘two’ >PMP *duha >Itbayat *duha*

b. PAn *Sadu ‘many, much’ >PMP *hadu >Itbayat *aru*

c. PAn *Salas ‘forest’ >PMP *halas >Itbayat *axas*

The morphological innovations of Proto Malayo-Polynesian are more complicated. They consist of innovations in pronouns and in verbal affixes.

A major set of innovations in pronouns involved a ‘politeness shift’. These changes have been reconstructed by Blust (1977) and in more detail by Ross (2002b). To the extent that they are relevant to the history of the Batanic languages, they are shown in (5).⁽¹⁵⁾ In English *you*, the second person plural pronoun, has replaced *thou*, its former singular counterpart, for reasons of politeness. Similarly, the PAn second person genitive singular *=Su ‘thy’⁽¹⁶⁾ was replaced in PMP by the corresponding PAn plural *=mu ‘your’, which became =*mu* in the Batanic languages.⁽¹⁷⁾ At about the same time plural *=mu was disambiguated from the singular (just as speakers of some English dialects use *you-s* or *y’all* for the plural) in one of two ways. It was either reinforced by the addition of the PAn second person free plural pronoun *iSu ‘thou’, giving PAn *=mu-iSu, which by regular sound change became

(15) The politeness shift also affected free pronouns, but this is not reflected in the Batanic languages, where free pronouns have undergone a further shift.

(16) ‘=’ indicates that the form was bound to the preceding word, cf. Itbayat *mangga=mu* ‘your mango’.

(17) Starosta (2002) claims that reflexes of *mu are found as the initial part of certain second-person pronouns in Formosan languages, and that the ‘politeness shift’ is therefore not a Proto Malayo-Polynesian innovation. Most of his examples are from Rukai, where the *m-initial forms probably reflect an old genitive marker, not the plural pronoun (Ross 2003).

(5)

second person genitive pronoun	Proto Austronesian	Proto Malayo-Polynesian	Itbayat
singular	*=Su	*=mu	=mu
plural	*=mu	*=mu-ihu *=nihu	=miyu =ñu

PMP *=mu-ihu, Itbayat =miyu, or it was replaced by PAn *ni-iSu 'of-thou', becoming PMP *=nihu, and =ñu in all Batanic languages.⁽¹⁸⁾

Proto Austronesian and Proto Malayo-Polynesian both had various sets of free-standing pronouns. Proto Austronesian also had a single set of bound pronouns (*i.e.* pronouns attached to the preceding word) which were used as both genitives and nominatives. This set continued in Proto Malayo-Polynesian, but only as genitives. A new set of nominative bound pronouns appeared (derived from earlier freestanding pronouns) (Ross 2002b:51). This set was defective: it lacked first and second person plural forms,⁽¹⁹⁾ and its third person forms are not reflected in many Malayo-Polynesian languages. The Batanic languages reflect the first and second person singular forms of the Proto Malayo-Polynesian nominative set, as shown in (6).

(6)

person and number	Proto Austronesian	Proto Malayo-Polynesian genitive	Proto Malayo-Polynesian nominative	Itbayat genitive	Itbayat nominative
1S	*=ku	*=ku	*=aku	=ku	=aku
2S	*=Su	*=mu	*=kaw	=mu	=ka

Both Proto Austronesian and Proto Malayo-Polynesian had complex and unusual verbal systems, and these systems are largely shared by their modern descendants in Taiwan and the Philippines. A description of these systems lies beyond the scope of this paper. However, it is significant that the Proto Malayo-Polynesian verbal system underwent innovations introducing complexities that are not present in Formosan systems, and the Batanic languages share in these complexities.

Among other things, Proto Malayo-Polynesian added to the system deriving verb roots

(18) It is unusual for a singular form to acquire plural function as part of a politeness shift, and this change remains unexplained.

(19) Except *=ta 1IP, which served as both genitive and nominative.

the prefixes *paN- ‘distributive’,⁽²⁰⁾ *paR- ‘durative, reciprocal’ and *paka- ‘aptative, potential’.⁽²¹⁾

The prefix *paN- is reflected as Yami (Imurud) *paN-* ‘augmentative’ (Shih 1997:76–77). Reid (1966) does not give the meanings of affixes, but records Ivatan *paN-* (his *pang-*) in his Verb Stem Classes 1–5. Tsuchida *et al.* (1989) record reflexes of *paN- for Yami (Iranumilek), Itbayat, Ivatan (both dialects) and Babuyan in their sentences 35 and 47.

The prefix *paR- is reflected as Yami (Imurud) *pi-* ‘reciprocal’ (Shih 1997:78–79). Reid (1966) records Ivatan *pay-* in his Verb Stem Classes 8–12. Tsuchida *et al.* (1989) record reflexes of *paR-: Yami (Imurud, Iranumilek) and Itbayat *pi-* and Ivatan (both dialects) and Babuyan *pay-*.⁽²²⁾

The prefix *paka- ‘aptative, potential’ is less widely reflected in the data. Yami (Imurud) has the reflex *paka-* ‘aptative’ (Shih 1997:82–83). This also occurs once in Tsuchida *et al.* (1989) in Yami (Iranumilek), sentence 43. Its absence from the data for the other languages is probably a result of the fact that aptative and potential meanings (‘can’, ‘be able’, ‘may (happen)’) are often not part of a linguist’s preliminary data collection.

2.1 First conclusions and more questions

We have seen above that the Batanic languages reflect the innovations which Malayo-Polynesian languages share, and are therefore one of the 20–25 western Malayo-Polynesian groups shown in Figure 2. This allows us to exclude the possibility that speakers of Proto Batanic arrived in the islands from Taiwan after the departure from there of speakers of Proto Malayo-Polynesian. Instead, we can be confident that Proto Batanic is a descendant of Proto Malayo-Polynesian. However, this still leaves us with two possible histories.

History 1 is that the more or less immediate ancestors of Proto Malayo-Polynesian speakers migrated out of Taiwan (around 4000 BC) to one or more of the islands where the Batanic languages are now spoken. Their speech underwent certain innovations and

(20) The reconstruction of *paN- and *paR- is briefly discussed by Ross (2002b:49–50). Blust (1999:68) and Starosta (2002:194–195) have claimed that *paN- is reflected in Formosan languages and are therefore not a Proto Malayo-Polynesian innovation. However, the idea that there is or has been a Formosan reflex of *paN- with the massive role of its Proto Malayo-Polynesian reflexes is clearly incorrect. Blust gives a single, possibly fossilised example from Puyuma. If a reflex of *paN- had played a role in a language ancestral to Puyuma anything like its Proto Malayo-Polynesian role, then it would have left numerous traces. Starosta gives examples from Amis, but my own research indicates that his morphological analysis is wrong and that these are not reflexes of *paN-.

(21) I am not aware that *paka- has been formally reconstructed, but its reflexes in Tagalog, Cebuano Bisayan and Binukid, for example, are self-evident.

(22) See Tsuchida *et al.* (1989), sentences 16, 43, 52, 85, 87. Note that in their sentences the prefixes written here as *paN-*, *pi/pay-* and *paka-* are usually modified by the presence of adjacent affixes or by the root-initial consonant.

became Proto Malayo-Polynesian. One or more groups of Proto Malayo-Polynesian speakers then migrated southward to Luzon, but leaving a Proto Malayo-Polynesian speaking community behind. The speech of this community underwent the innovations exclusively shared by the Batanic languages, *i.e.* it became Proto Batanic. This means that if we draw a more detailed family tree of Malayo-Polynesian, the bifurcation beneath the Proto Malayo-Polynesian node is into Proto Batanic and an interstage language ancestral to all Malayo-Polynesian languages except Batanic.

History 2 offers no hypothesis about exactly where Proto Malayo-Polynesian was spoken, but infers that it diversified into various speech communities in northern Luzon. At some later date, members of one of these communities migrated north into the Batanes Islands, and their speech became the ancestor of Proto Batanic.

If it can be demonstrated that Batanic belongs in some subgroup subordinate to Malayo-Polynesian, then we will have evidence in favour of History 2. This is the topic of the investigation reported in §3. If, however, Batanic cannot be shown to have a link with another Philippine group, this does not exclude History 2: it could easily be the case that its relative(s) on Luzon have become extinct. Thus an analysis of the subgrouping relations of Batanic cannot tell us with certainty whether History 1 is true. I return to this problem in §4.

Is there evidence in favour of History 2? To answer this question, we need to understand something of the relationships among the languages of the northern Philippines. Unfortunately, there is really no definitive account of these, and this is a problem which we must now address.

3 PHILIPPINE RELATIONSHIPS AND PROTO BATANIC

We might expect the linguistic history of the Philippines roughly to resemble that of Taiwan, as follows:

- a. Dialect variation in Proto Malayo-Polynesian would eventually have led to loss of mutual intelligibility between dialects and to the emergence of separate languages, some of which would be ancestral to groups of today's Philippine languages;
- b. Just as there was no 'Proto Formosan', so there would also have been no 'Proto Philippine'.
- c. Just as a group of speakers migrating out of Taiwan became the Proto Malayo-Polynesian speech community, so we might expect the speech of one or more groups migrating to Borneo and/or Sulawesi to have become reconstructable interstages ancestral to the extra-Philippine Malayo-Polynesian languages.

Expectation (c) has not yet been met. This may be the result of a very rapid dispersal of emigrating speakers that left no time for innovations which would allow the identification

of distinct interstages. Or it may be because of insufficiently detailed historical linguistic research in western Indonesia: a survey of recent work suggests that an interstage language ancestral to all Sulawesi languages and to languages outside Sulawesi may yet be reconstructable (Ross 2002a:459–470).

Expectation (a) is satisfied: basing himself on scholarly consensus, Blust (1991) lists fifteen ‘microgroups’ of languages (one of them is Batanic) spoken in the Philippines region (which includes the Philippines but extends north to Lányǔ and south to the northern arm of Sulawesi). But he also records a fair amount of scholarly disagreement about the shape of the family tree of which these microgroups would be the terminal nodes.

Expectation (b), no ‘Proto Philippine’, is somewhat problematic. Until Reid (1982) there was a long accepted convention that the languages of the Philippines were indeed descended from a ‘Proto Philippine’, a descendant of Proto Malayo-Polynesian. This was based on a perception that the languages of the Philippines region appeared too like each other and too unlike languages immediately to their south for them **not** to be descended from a unique ancestor. Reid recognises, however, that only exclusively shared innovations, and not similarity, are sufficient grounds to posit a unique ancestor. He points out that the grammatical features common to Philippine languages are widely enough reflected outside the Philippines to reconstruct them for Proto Malayo-Polynesian (Reid calls it ‘Proto Extra-Formosan’). That is, they are not exclusively shared innovations. He shows that some of the phonological innovations allegedly reflected in all languages of the Philippines region were fictitious, because they were based on inaccurate reconstruction of Proto Malayo-Polynesian phonology. Other phonological innovations, Reid reminds us, are shared by languages outside the region and are therefore not evidence for ‘Proto Philippine’.

It would seem that Reid has effectively demolished the case for ‘Proto Philippine’. This is not quite the end of the story, however. Zorc (1986) is a counterblast to Reid, and argues that there are enough **lexical** innovations shared by Philippines region languages to justify the reconstruction of Proto Philippine after all. Zorc reconstructs 98 words which he believes are found only in Philippines region languages.⁽²³⁾ Ninety-eight lexical innovations is a substantial number, and Zorc says that he has eliminated all instances that he can recognise as borrowing from one Philippine language into another. He admits that some of these items may yet be found in extra-Philippine languages (and will then cease to be exclusively shared Philippines innovations), but thinks it unlikely that more than a small minority of his putative innovations will be invalidated in this way.

What are we to do with these facts? Blust (1991:77) is convinced by the lexical innovations amassed by Zorc, and suggests (2000:108–109) that sometime after the dispersal of

(23) Twenty-three of them are widely distributed across the region, and the other 75 are reflected in widely separated parts of the region.

Proto Malayo-Polynesian speakers across the Philippines and after the diversification of their speech into various languages, one Malayo-Polynesian language expanded across the Philippine region, extinguishing earlier linguistic diversity. This language was Proto Philippine.

I see a problem here. If this putative Proto Philippine had had time to undergo so many lexical innovations, then why are its reconstructible phonological and grammatical (specifically, its verbal) systems identical to those of Proto Malayo-Polynesian? Why have they not also undergone innovations? Is there perhaps a better explanation for Zorc's 98 apparent lexical innovations, one that does not necessitate positing a Proto Philippine interstage? It seems to me that there are two such explanations, and they are not mutually exclusive.

The first explanation is that some of Zorc's vocabulary items have been retained from Proto Malayo-Polynesian but lost in extra-Philippine Malayo-Polynesian languages. If, as I suggested above, Malayo-Polynesian languages outside the Philippines are descended from only a very small number of speech communities emigrating out of the southern Philippines, then it is quite plausible that these communities had lost certain vocabulary items, leaving them extant only in languages remaining in the Philippines.

The other explanation is that some of these words were innovated somewhere in the Malayo-Polynesian dialect network in the Philippines soon after the separation of dialects ancestral to extra-Philippine Malayo-Polynesian languages, and consequently gained currency across the network as a result of contact.

Under neither explanation would shared vocabulary items reflect the erstwhile existence of Proto Philippine, and, in the absence of phonological and grammatical innovations, I remain skeptical that a Proto Philippine ever existed.

Blust (1991:104) observes that 'the Philippine archipelago as a whole shows much less linguistic diversity than one would expect for a region that must have been settled very early in the history of the Austronesian expansions.' It is on this basis that he infers an extinction of diversity resulting from the expansion of Proto Philippine. However, there is no reason to think that languages necessarily continue to diversify over time. Whether they do so or not is very much an outcome of the structure of social networks. In Taiwan, it seems, there was diversification. But where communities maintain quite intense social and economic relations and speakers are bilingual in the languages of neighbouring communities, there will be a strong tendency for similarities between neighbouring languages to be retained. This provides an alternative explanation both for the conservatism of Philippine region languages and for their similarities.

If the only common ancestor of Philippine languages was Proto Malayo-Polynesian, then we are not immediately any the wiser about the provenance of the Batanic languages than we were at the end of §2.1: either History 1 or History 2 is possible. We must therefore turn to the internal subgrouping of northern Philippines languages.

All recent attempts to organise the Philippine microgroups listed by Blust (1991) into larger groupings agree that there is a major division between north and south Luzon, shown in Figure 1. They also agree that there are three distinct groups north of this line, labelled Batanic, Cordilleran and Central Luzon in Figure 1.

There are no shared innovations other than lexicon to suggest that Batanic, Cordilleran and Central Luzon form a larger grouping, and Reid (1982) does not group any of them together. Zorc (1986) places them in a Northern Philippine group, along with the North Mangyan group on the island of Mindoro. He divides his Northern Philippine group into two subgroups on lexical grounds: the first consists of Batanic, Central Luzon and North Mangyan, and the second is Cordilleran.⁽²⁴⁾ Reid (pers. comm.) rejects the Northern Philippine grouping but accepts the two subgroups as each is characterised by shared phonological innovations. The first group is characterised by an innovation whereby PMP *R > y. The Cordilleran languages as a whole share few innovations other, perhaps, than shared lexicon. PMP *h (from PAn *S, *h) has become zero, and so has PMP word-final *-q.

Blust (1991:79–80) accepts that Central Luzon and North Mangyan may yet prove to be a larger genetic unit, but rejects a unit which would also include Batanic. He rejects the PMP *R > y innovation here as subgrouping evidence on the grounds that the same change has also occurred in scattered languages and groups outside the Philippines, implying that its presence in two Philippine microgroups could well be due to independent parallel innovation (Blust 1991:106).⁽²⁵⁾

An argument similar to the one above against the lexical evidence for Proto Philippine can be applied to the lexical evidence for a possible Proto Batanic/Central Luzon. Given the geographic separation of Batanic and Central Luzon, it is likely that uniquely shared lexical items reflect retentions from Proto Malayo-Polynesian which were lost in Proto Cordilleran and in the interstage languages ancestral to languages south of the Luzon divide. Indeed, this seems considerably more probable than that they reflect a unique common ancestor.

No morphological evidence of a link between Batanic and Central Luzon has been offered, and inspection of the systems of pronouns, articles and other functors listed by McFarland (1977:11–23) indicates none.

3.1 A conclusion

Blust (1991:79–80) writes, 'It may well prove true . . . that Central Luzon and North Mangyan form a larger genetic unit, but I am aware of no convincing evidence that Bashii

(24) The relevant data have not been published.

(25) Interestingly, he accepts PMP *R > g as evidence of his Greater Central Philippine grouping, apparently because it is a rarer—though not unique—phonological change and because it is backed by 94 putative lexical innovations (Blust 1991).

[=Batanic] is more closely related to it than to any other Philippine microgroup' (my brackets). If the only common ancestor of Philippine languages was Proto Malayo-Polynesian, and Batanic has no special link with any other Philippine group, then, as I observed in §2.1, either History 1 or History 2 remains possible. The fact that Batanic has no link with another Philippine group does not exclude History 2, the possibility that it is descended from a returnee group from Luzon, as one or more related languages on Luzon may have become extinct.

4 THE CONSERVATISM OF PROTO BATANIC

This may appear at first sight to be a rather disappointing conclusion. However, it is significant that History 1 remains a possibility: had we found that Batanic formed a larger grouping with a language or languages of Luzon, then History 1 would have been eliminated in favour of History 2.

There is a possible way out of this impasse. It is provided by the principle that, other things being equal, the speech of a community that remains in the same location will be subject to fewer innovations than the speech of a community which changes location. This principle is not accepted by all historical linguists, but it is based on a logic which says that a community which is stationary is less likely to undergo social change than a community which moves. Whilst there is a tendency for any language to change, the rate of change is increased by social selection, *i.e.* the adoption of an innovative feature as a social marker (Nettle 1999:29–30, 48–51). A community which migrates is likely to undergo changes in social structure, and its speech is therefore likely to undergo increased social selection. Language change is also likely to occur as a result of contact with new groups of people and as a result of a change of environment and of adapting to it.

Conservatism is difficult to measure, but a comparison of the Batanic languages with reconstructed Proto Malayo-Polynesian suggests that they have changed far less than the languages of northern Luzon. As the examples above indicate, other than the change of PMP *R to Proto Batanic *y, very little phonological change has occurred. I have made no count of Batanic lexical items, but the retention rate from Proto Malayo-Polynesian appears on inspection to be high. The verbal systems of Yami and Ivatan appear very like the one which is reconstructable for Proto Malayo-Polynesian.

One fairly well researched area of comparative Philippines grammar is that of the case-marking articles (alias 'noun phrase markers', alias 'construction markers') which occur at the beginning of a noun phrase. Reid (1978) showed that two 'grades' of article are reconstructable with common noun phrases. The 'grade' is marked by the vowel: *-i or *-u. Reid's and my own unpublished researches indicate that a third grade in *-a should also be reconstructed for Proto Malayo-Polynesian. The meaning contrasts between the three grades apparently had to do with a number of factors, among them spatial relations (*e.g.* whether the referent of the noun phrase is near the speaker, near the hearer or near neither, and whether it is present or absent), time (past, present and future), and specificity and definite-

ness (i.e. whether the referent can be identified by the speaker, by the speaker and the hearer, or by neither).⁽²⁶⁾ The reconstructable Proto Malayo-Polynesian forms are listed in (7).

(7)

Proto Malayo-Polynesian				
	Nominative	Genitive	Oblique	Locative
*-i grade	*i	*ni	*si	*di, *i
*-a grade	*a	*na	*ta, *sa	*da
*-u grade	*u	*nu	*tu, *su	*du (?)

The Malayo-Polynesian languages which retain a set of case-marking articles at all are mainly located in the Philippines region. Most of these, however, have lost the grade contrast, like Ivatan in (8), which retains only the *u*-grade. If one tabulates case-marking articles from across the region, it is fairly clear that the forms in (7) can be reconstructed, yet most languages maintain a random mixture of grades or at most just two grades.

Itbayat is one of the very few languages which retains a grade contrast, as (8) shows. The *a*-grade forms have disappeared, but the *i*- and *u*-grade forms remain.⁽²⁷⁾

(8)

		Nominative	Genitive	Oblique	Locative
Itbayat	proximal	<i>i</i>	<i>ni</i>	<i>si</i>	<i>di</i>
	distal	<i>u</i>	<i>nu</i>	<i>su</i>	<i>du</i>
Ivatan		<i>u</i>	<i>nu</i>	<i>su</i>	<i>du</i>

This makes Itbayat one of the most conservative of Malayo-Polynesian languages with regard to this grammatical feature. There is one context, however, in which retention of the grades is even more conservative, namely when a case-marking article is prefixed to a demonstrative ('this', 'that'). Here, as (9) shows, all three grades are maintained, both in Itbayat and in Ivasay Ivatan.⁽²⁸⁾

(26) In Ross (2002b) I glossed the *-i, *-a and *-u grades as 'default', 'present' and 'absent' respectively. Lawrence Reid (pers. comm.) rightly says that this is too simple, as meaning distinctions between the members of such sets are never this simple.

(27) Tsuchida *et al.* (1987:22) record the Ivatan locative common form as *di*, but Reid (pers. comm.) points out that this is the locative personal form.

(28) The Itbayat nominative near-addressee form *uri* is perhaps a reduction of expected **a-uri*.

(9)

	Itbayat demonstratives			Ivatan demonstratives	
	Free	Nominative	Locative	Nominative/ Predicate	Locative
Near speaker	<i>ni-ya</i>	<i>i-ya</i>	<i>di-ya</i>	<i>ni-aya'</i>	<i>di-a'</i>
Near addressee	<i>na-wi</i>	<i>uri</i>	<i>da-wi</i>	<i>na-wri'</i>	<i>da-w</i>
Away from both	<i>nu-urihay</i>	<i>u-urihay</i>	<i>du-urihay</i>	<i>nu mian uri'</i>	<i>du-nguriaw</i>

I know of only one other language which maintains this same three-way contrast (others do retain a two-way contrast, however, e.g. Tagalog *di-to* 'here' vs. *do-on* 'there'). This is Mamanwa, a Negrito language of East Mindanao in the southern Philippines.⁽²⁹⁾ The most reasonable way of accounting for the Itbayat and Mamanwa sets is to infer that both are highly conservative and have retained a more complex Proto Malayo-Polynesian system that has been simplified in other languages.

If the Batanic languages are indeed very conservative (and this needs to be confirmed more formally), then there is a greater likelihood that History 1 is correct. That is, Proto Malayo-Polynesian came into being on one or more of the islands where the Batanic languages are now spoken. One or more groups of Proto Malayo-Polynesian speakers migrated southward to Luzon, leaving behind a Proto Malayo-Polynesian speaking community whose speech became Proto Batanic.

The evidence for this conclusion is circumstantial, and does not have the same status as subgrouping evidence based on shared innovations. Logically, however, if History 1 is true, there can be no subgrouping evidence based on shared innovations.

4.1 The similarity of the Batanic languages

The apparent conservatism of the Batanic languages raises another issue. As noted earlier, the Batanic languages are remarkably similar to each other. In the case of Babuyan, we know that this is because ancestors of its present-day speakers came from Ivatan-speaking communities about a century ago. There are greater differences, at least in phraseology, between Yami and the other three languages, but these are perhaps the quite recent outcome of the international boundary that runs between them. Indeed, it is often assumed that similarity between the languages of a group indicates that the protolanguage has only very recently broken up. If this assumption were true of all the Batanic languages, then it would also raise the question, On which island was Proto Batanic spoken?

(29) Reid (pers. comm.) points out that Negrito languages tend to retain older features.

The available data allow no answer to this question, but perhaps it should not be answered, as there is another possible explanation of Batanic homogeneity. If communities on the islands where Batanic has been spoken have remained in continuous and intensive contact, then linguistic unity may well have been maintained over a very long period of history. The language would have changed slowly, but the speech conventions of the various island communities would have remained more or less in sync with one another, maintaining their similarity.

5 WHENCE THE MALAYO-POLYNESIANS?

If indeed the Batanic languages do reflect a very early offshoot of Malayo-Polynesian, then it becomes relevant to ask where their immediate ancestors, the speakers of Proto Malayo-Polynesian, came from.

Some confusion concerning the Proto Malayo-Polynesian speaking community has appeared in recent literature. Starosta (2002) cites a number of scholars who have described Malayo-Polynesian as a ‘first-order subgroup’ of Austronesian, and presents a number of Austronesian family tree diagrams by these scholars in which, as in Figure 2 above, Proto Malayo-Polynesian is shown as branching off directly from the Proto Austronesian node. Starosta is at pains to show that Malayo-Polynesian was not a ‘first-order subgroup’ of Austronesian because Proto Austronesian must have diversified into a dialect chain and/or undergone a recursive process of language splitting before Proto Malayo-Polynesian came into being. This would mean, according to Starosta, that Proto Malayo-Polynesian has a position two or more nodes further down the tree than I have placed it in Figure 2, for example, and that it is therefore not a first-order subgroup.⁽³⁰⁾

I am not aware of any scholar who would disagree with this general characterisation of the position of Proto Malayo-Polynesian,⁽³¹⁾ and it seems to me that Starosta uses ‘first-order subgroup’ in a much stricter sense than most historical linguists, and that he reads a degree of precision into family tree diagrams that the scholars who draw them do not intend. Proto Malayo-Polynesian is shown as branching off from the Proto Austronesian node in Figure 2 (and in many similar figures in the literature) because I am making no attempt to depict the details of linguistic diversification between the initial diversification of Proto Austronesian and the emigration from Taiwan of the Proto Malayo-Polynesian speaking community. It does not represent a commitment to the idea that Proto Malayo-Polynesian was spawned by the initial break-up of the Proto Austronesian speaking community into a first network of dialects of languages.

(30) Starosta also tries to show that some of the innovations shared by Malayo-Polynesian languages are also shared by one or more Formosan languages. I have indicated in two earlier footnotes that two of his claims are ill-founded. I will refer below to a third.

(31) Unless s/he rejects the very notion of Proto Malayo-Polynesian.

If, then, the Proto Malayo-Polynesian speech community left Taiwan after a period of Austronesian diversification within Taiwan, can we say where it came from? In other words, can we say that it is more closely related to certain Formosan languages than to others?

My view is that we probably cannot. Whatever the closest relative of Proto Malayo-Polynesian in Taiwan may have been, that information is no longer recoverable with any certainty.⁽³²⁾

Blust (1999) provides a new classification of Formosan languages into nine groups based on phonological innovations. None of these groups displays all the three innovations shared by Malayo-Polynesian languages as listed in (1). Bunun, a southern language which forms a single-member group, as well as the members of the northern branch of Blust's East Formosan group (Basay – Trobiawan, now extinct, and Kavalan), do share the first two Proto Malayo-Polynesian innovations in (1), namely the merger of PAN *t and *C and the merger of *L and *n. However, no Formosan language shares the PMP *S > h innovation.

The phonological innovations of Proto Malayo-Polynesian are compared in (10) with

(10)

	East Formosan					
	PMP	Bunun	North	Amis	Siraya	Kanakanabu
*t/C merger	yes	yes	yes	yes	yes	no
L/n merger	yes	yes	yes	no	no	yes
S > glottal spirant	yes	no	no	no	no	no
j/n merger	no	no	yes	yes	yes	no
q lost	no	no	yes	no	yes	no
q > ?	no	no	no	yes	no	no
l/R merger	no	no	no	yes	no	no
d/s merger	no	no	no	no	yes	no
S/R merger	no	no	no	no	yes	no
k/g merger	no	yes	no	no	no	yes
j/l merger	no	yes	no	no	no	no
S/s merger	no	yes	no	no	no	yes
r/R merger	no	yes	no	no	no	yes

(32) Sagart (2004) offers an alternative view which came to my attention after the completion of the present paper.

those of the three branches of East Formosan (northern, Amis, Siraya), Bunun, and Kanakanabu, as these are the Formosan groups and languages that share at least one phonological innovation with Proto Malayo-Polynesian.

Bunun has four more mergers which are not shared by Proto Malayo-Polynesian. If Bunun were the closest relative of Proto Malayo-Polynesian, then this would mean that their unique common ancestor would have undergone the mergers of PAn *t and *C and of *L and *n before the departure of the Malayo-Polynesians.

The northern branch of East Formosan shares the odd *j/n merger with Amis and Siraya, and this is Blust's strong reason for uniting them in his East Formosan group. Amis and Siraya, on the other hand, do not share in the *L/n merger. If East Formosan were the closest relative of Proto Malayo-Polynesian, then their unique common ancestor would have undergone only the merger of *t and *C before the departure of the Malayo-Polynesians, since Proto Malayo-Polynesian does not share in the *j/n merger which characterises East Formosan, and the *L/n merger has not occurred in Amis or Siraya.⁽³³⁾

Bunun is a better candidate for the closest relative of Proto Malayo-Polynesian, with two shared innovations. East Formosan is a very weak candidate, as the only shared innovation is the *t/C merger, and this, as we have just noted, has also occurred in Bunun. That is, it must have occurred independently on at least two occasions in the early history of Austronesian languages, and is therefore poor subgrouping evidence.

Kanakanabu underwent the *L/n merger. This, however, is even weaker subgrouping evidence, as it seems to have occurred independently at least three times.

None of these subgrouping hypotheses has merit as it stands. It is just possible that further research will reveal shared non-phonological innovations, but this seems unlikely.

Harvey (1982) has argued that the closest Formosan relative of Proto Malayo-Polynesian is Amis, spoken along much of the east coast of Taiwan. He gives three putative innovations shared by Amis and Proto Malayo-Polynesian:

- (11) a. PAn *t and C have merged.
- b. Both have innovatory pronoun forms reflecting *ka-.
- c. Both reflect a third-person pronoun *si-da.

None of these innovations withstands scrutiny. Innovation (11a) is indeed shared, but

(33) The northern branch of East Formosan underwent a *j/n/L merger.

of scant value, as we observed above. Innovation (11b) rests on the assumption that pronoun forms reflecting *ka- are innovatory, but they aren't. Polite pronouns in *ka- are reflected in certain Formosan languages and are therefore not an innovation in Proto Malayo-Polynesian. In any case, the *k*-initial pronouns in Amis have a different origin: they are nominative pronouns, and *k*- is the nominative marker (contrasting with genitive *n*- and accusative *t*-). As for innovation (11c), the proposed cognates are Amis *sira* 's/he' and Proto Malayo-Polynesian *sida 'they'. Their meaning difference makes them an unlikely cognate pair.

Reid (1982:212) points out that Tagalog *qamihan*, Ilokano *qamian* 'north wind' may reflect PAN *amiS-an 'the place of the Amis'. However, the putative fact that Proto Malayo-Polynesian speakers associated the north with the Amis homeland does not necessarily mean that they came from there. Indeed the word seems to have meant 'north' before Proto Malayo-Polynesian speakers left Taiwan, as it is reflected in Formosan languages (Blust In progress): Atayal *qmis-an* 'winter', Puyuma *Hami* 'north; year, age', Paiwan (northeastern dialect) *qamis* 'north', Amis *?amis* 'north'.

The case for Amis as the closest Formosan relative of Proto Malayo-Polynesian is thus non-existent, and we are left with no obvious alternative candidate.

REFERENCES

Blust, Robert A.

- 1977 The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *University of Hawaii Working Papers in Linguistics*, 9/2:1-15.
- 1990 Patterns of sound change in the Austronesian languages. In Philip Baldi, ed., *Linguistic change and reconstruction methodology*, 231-263. Berlin: Mouton de Gruyter.
- 1991 The greater central Philippines hypothesis. *Oceanic Linguistics*, 30:73-129.
- 1999 Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds., *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 31-94. Taipei: Academia Sinica. (Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica 1.)
- 2000 Chamorro historical phonology. *Oceanic Linguistics*, 39:83-122.
- In progress. Austronesian comparative dictionary. Computer files.

Harvey, Mark

- 1982 Subgroups in Austronesian. In Amran Halim, Lois Carrington, and S. A. Wurm, eds., *Papers from the Third International Conference on Austronesian linguistics*, 47-99. Canberra: Pacific Linguistics. (Pacific Linguistics C-75.)

Hidalgo, Cesar

- 1996 *The making of the ivatans: The cultural history of batanes*. Pasig, Metro Manila: Cognita TRC.

Maree, Rundell

- 1982 *Ibatan cultural origins and social organization*. M.A. thesis, University of Texas at Arlington.

McFarland, Curtis D.

- 1977 *Northern Philippine linguistic geography*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa. (Studies of Languages and Cultures of Asia and Africa, Monograph Series 9.)

Nettle, Daniel

- 1999 *Linguistic diversity*. Oxford: Oxford University Press.

Pawley, Andrew

- 2002 The Austronesian dispersal: Languages, technologies and people. In Peter Bellwood and Colin Renfrew, eds., *Examining the farming/language dispersal hypothesis*, 251–273. Cambridge: MacDonal Institute for Archaeological Research, University of Cambridge.

Reid, Lawrence A.

- 1966 *An Ivatan syntax*. Honolulu: University of Hawaii Press. (Oceanic Linguistics Special Publication No. 2.)
- 1978 Problems in the reconstruction of Proto-Philippine construction markers. In S. A. Wurm and Lois Carrington, eds., *Second international conference on Austronesian linguistics: Proceedings*, 33–66. Canberra: Pacific Linguistics. (Pacific Linguistics C-61.)
- 1982 The demise of Proto-Philippines. In Amran Halim, Lois Carrington, and S. A. Wurm, eds., *Papers from the Third International Conference on Austronesian linguistics*, 201–216. Canberra: Pacific Linguistics. (Pacific Linguistics C-75.)

Ross, Malcolm

- 1992 The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics*, 31:23–64.
- 1995 Some current issues in Austronesian linguistics. In D. T. Tryon, ed., *Comparative Austronesian dictionary*, 45–120. Berlin: Mouton de Gruyter.
- 2002a Final words: Research themes in the history and typology of western Austronesian languages. In Fay Wouk and Malcolm Ross, eds., *The history and typology of western Austronesian voice systems*, 451–474. Canberra: Pacific Linguistics.
- 2002b The history and transitivity of western Austronesian voice and voice-marking. In Fay Wouk and Malcolm Ross, eds., *The history and typology of western Austronesian voice systems*, 17–62. Canberra: Pacific Linguistics.
- 2003 Notes on reconstructing Proto Rukai case-marking. Unpublished ms.

Sagart, Laurent

- 2004 The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic*

Linguistics, 43:411-444.

Scheerer, Otto

1926 Batan texts with notes. *Philippine Journal of Science*, 31:301-341.

Shih, Louise Yu-mien

1997 *Yami morphology*. M.A. thesis, Providence University, Taijung, Taiwan.

Starosta, Stanley

2002 The rise and fall and rise and fall of proto malayo-polynesian. In Robert S. Bauer, ed., *Collected papers on southeast Asian and Pacific languages, 185-203*. Canberra: Pacific Linguistics.

Tsuchida, Shigeru, Ernesto Constantino, Yukihiro Yamada, and Tsunekazu Moriguchi

1989 *Batanic languages: Lists of sentences for grammatical features*. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.

Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi

1987 *Lists of selected words of Batanic languages*. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.

Zorc, R. David

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, 147-173*. Canberra: Pacific Linguistics.
