

Primary subgroups of Austronesian: a consideration of some hypotheses*

Malcolm Ross

The Australian National University

Malcolm.Ross@anu.edu.au

January 24, 2009

Draft only. Please do not cite without the author's agreement.

1 Introduction

Blust (1977) proposed that Austronesian languages fall into a number of primary subgroups within mainland Taiwan and just one subgroup, Malayo-Polynesian, embracing all Austronesian languages outside mainland Taiwan. The 'Out of Taiwan' hypothesis has gained in support since 1977, as further shared Malayo-Polynesian innovations have been identified. It follows from the hypothesis that if we are to learn more about the earliest history of Austronesian and its primary subgroups, it is mostly to Taiwan that our attention must turn.

There have been numerous proposals about the subgrouping of the Formosan languages: Blust (1999b:37–40) summarises seventeen which predated an eighteenth presented in that paper. Attention will here be limited to three proposals: Ho (1998), Blust (1999b) and Sagart (2004). These three are chosen not because they are recent, but because they represent applications of the classical comparative method of linguistics which seeks to subgroup languages on the basis of shared innovations. Most earlier proposals incorporated at least bits of other methodologies, and Blust (1999b) provides a critique of these which need not be repeated here.

The point of interest in these three papers is that they each employ the comparative method, yet they arrive at what appear to be very different subgroupings. In theory, at least, if the comparative method is thoroughly and properly applied to the available data, it should lead to only one set of results! It emerges that the different subgroupings arise out of differences in data, differences in the analysis of shared innovations, and differences in the evaluation of an innovation's significance. This last plays a major role in the strikingly different analyses of Blust and Sagart. Blust sees shared phonological

*I am very much indebted to Stacy Fang-ching Teng, who read and summarised Chinese-language articles for me. I am also grateful to Robert Blust and Peter Trudgill for answering my questions during the preparation of this paper.

innovations as having major significance, while Sagart discounts them almost entirely. Section 2.3 is devoted to this matter.

My approach here is to present each of the three subgroupings with its supporting innovations, and to offer an assessment of the significance of each innovation and sometimes of the analysis which underlies an innovation, proffering a revised (and flatter) version of that subgrouping based on this assessment. In §5 I compare the three revised subgroupings and show that they can in large measure be reconciled with one another, although important issues remain to be resolved by future research.

Although the publication of Ho (1998) predates that of Blust (1999a) by one year, it appears that neither scholar had seen the other's work at the time of writing. For convenience of presentation I discuss Blust's subgrouping first, then Ho's, before moving on to Sagart's (2004) work.

2 Blust's subgrouping

2.1 Review

The phonological innovations in Formosan languages listed by Blust (1999b:44–45) are set out in Table 1.¹ The format of that table seeks to show graphically which languages show which innovations. A bullet (●) indicates a categorical innovation, a plus sign (+) a partial innovation, i.e. one that does not occur in every candidate lexical item, and a question mark (?) indicates that one of the phonemes involved in the merger is not attested in the language.

I have dismantled two subtypes of Blust's mergers. First, he shows some changes as mergers with zero, e.g. in the Basay variety of Ketagalan, $*q/\emptyset$. I have preferred to show these more conventionally as deletions, i.e. $*q > \emptyset$. This is particularly relevant in instances where he shows two or more phonemes merging with zero, e.g. in Taokas, $*p/k/j/\emptyset$. There is no reason to think that the three deletions encapsulated herein are connected, and so I show them as $*p > \emptyset$, $*k > \emptyset$ and $*j > \emptyset$, listing them with other lenitions at the top of the table.

Secondly, certain other mergers involve three or four phonemes. For example, Blust has Kavalan $*j/n/L$ merging. I prefer to treat this as three mergers, $*j/*n$, $*j/*L$ and $*n/*L$ (with no immediate assumption about the sequence of mergers) as this makes it easier to compare innovations across languages and to identify shared innovations.² By implication, Blust does the same, as he identifies $*j/*n$ as the major innovation diagnostic of the East Formosan subgroup. In a few cases I retain a merger of more than two phonemes, e.g. Hoanya $*S/*C/*L$. The $*S/*C$ merger occurs in more than one language, and is shown separately (and as occurring in Hoanya). The $*S/*L$ and $*C/*L$ mergers do not occur in languages other than Hoanya, however, and it is thus more economic to leave them combined as the $*S/*C/*L$ merger.

In the list of Puyuma innovations (Blust 1999b:44), a $*d/z/j$ partial merger is shown. This requires special attention. PAN $*j$ is consistently reflected as $\backslash d$ in Puyuma, whilst the sound correspondence table (p43) shows $*d$ and $*z$ merging as Puyuma $\backslash z$ and $\backslash d$. However, in a detailed discussion of borrowing between the two languages (pp47–51) Blust concludes that the inherited reflex of PAN $*d$ and $*z$ is Puyuma $\backslash z$, whilst $\backslash d$ occurs in borrowings. This being so, I treat this as a $*d/*z$ merger, excluding $*j$ from it.

Blust omits from his list of innovations certain mergers that are common in his sound correspondence table on the grounds that they are non-diagnostic. He names $*d/z$ and $*\tilde{n}/L$. I have included these in Table 1 for the sake of consistency (cf the previous paragraph). However, no reflexes of PAN $*\tilde{n}$, $*g$ and $*r$ are available in a number of languages, and so I have marked $*\tilde{n}/L$ and other mergers involving one of these phonemes with a question mark in these languages.

¹ I have omitted Kulon, as Blust says that the data are really too sparse to be reliable.

² Blust writes N where I prefer L, as this was probably a dental lateral (Ross 1992).

Table 1: Phonological innovations noted by Blust (1999) in Formosan languages

	Sai Paz	Ata	Tha	Tao	Bab	Pap	Hoa	Tso Kan Saa	Bun	Pai	Puy	Ruk	Sir	Ami	Bas	Tro	Kav	PMP	
*p > ∅				+	+	+													
*S > *h																		•	
*j > ∅ ¹				+				•	+	+			+						
*l > ∅									•										
*R > ∅										+					+				
*q > ?	•													•					
*q > ∅	•					•	•	•	•				•	•	•	•	•		
*k > ?								+											
*k > ∅				+	•	•													
*-w/y > ∅			+	•	•	•	•												
*C > s	•	•		•		•	•												
*t/*C									•					•	•	•	•	•	•
*t/*s ²			•	•	•?	•	•												
*t/*s/*L/*z ³			•																
-t/-j	•																		
*d/*s									•					•					
*d/*S									•										
*d/*j						•													
*d/*z	•	•	•	•			•	•	•	•		•		•	•		•		
*d/*z/*R															•				
*j/*s ⁴	•																		
*j/*n (>n)														•	•	•	•	•	
*j/*L								+							•	•	•		
*j/*ñ	?		?	?	?	?	?	+				?		?	?				
*j/*l							•												
*S/*R														•					
*S/*s								+	•	•	•							•	
*S/*C	•						•												
*S/*C/*L							•												
*s/*l															•	•			
*s/*l/*z																•			
*z/*j/*d?						•													
*z/*s/*S									•										
*r/*R	?	?	?	?	?	?	?	•	•	•	•	?	•	?	?	?	?	?	
*l/*r	?	?	?	?	?	?	?	•			?		?	?	?	?	?	•	
*l/*R	•							•						•					
*n/*L			•					+	•		•				•	•	•		
*ñ/*L	•	?	•	•	?	?	?	+	•	•	•	•	?	•	•	?	?	•	
*n/*ŋ			•	•	•	•													
*ñ/*ŋ	?			?	?	?	?		•			?		?	?				
*L > k								+											
*k/*g (>k)	?	?	?	?	?	?	?	+	•	•	•			?	?	?	?	?	
*j/*g	?	?	?	?	?	?	?						+	?	?	?	?	?	

¹ In Kananavu and Saaroa *j > ∅ preceding or following *i.
² Non-finally, except in Thao, where it is also reflected finally.
³ > /t/, as also does *j when it is not deleted.
⁴ Non-finally only.

Table 2: Lexical innovations noted by Blust (1999) in East Formosan languages

		Ami	Bas	Tro	Kav
*lalan	< *zalan ‘road’	•	•	•	
*nanum	< *daLum ‘water’	•	•		
*sisu	< *susu ‘breast’		•	•	•
*jaŋan ^l	< *ŋajan ‘name’	•?		•	•
*bajuR	< *baRuj ‘dove’	•			•

^l The Amis form is \ŋaŋan, via the operation of assimilation either on *naŋan, the expected reflex of *jaŋan, or on *ŋanan, the expected reflex of *ŋajan.

Blust also notes some idiosyncratic lexical innovations in East Formosan languages. These are shown in Table 2. The idiosyncratic forms reflected in the innovating languages shown are in PAN orthography together with the ‘normal’ PAN form.

Blust’s subgroups are separated by vertical lines in Table 1. From left to right they are (language names in italics):

- (1) a. Northwest Formosan (\Saisiyat, Pazeh)
- b. Atayalic (the dialects of \Atayal and \Seediq)
- c. Western Plains
 - i. Thao
 - ii. Central Western Plains (\Taokas, Favorlang–Babuza, Papora and \Hoanya)
- d. Tsouic (\Tsou, Kanakanavu and \Saaroa)
- e. Rukaic
- f. \Bunun
- g. \Paiwan
- h. \Puyuma
- i. East Formosan
 - i. \Siraya
 - ii. \Amis
 - iii. Northeast Formosan (\Ketagalan [\Basay and \Trobiawan varieties] and \Kavalan)
- j. Malayo-Polynesian

As I noted above, Table 1 includes innovations which are not diagnostic for subgrouping. Table 3 is a reduced version of Table 1 with non-diagnostic innovations removed. An innovation may be non-diagnostic either because it occurs only in one or two languages or because it is very common (and therefore as likely to reflect independent parallel innovation as it is to reflect shared inheritance). The innovations omitted from Table 3 are the following:

- (2) a. Reflected in just one language: *l > 0, *t/*s/*L/*z, *-t/*-j, *d/*S, *d/*j, *d/*z, *d/*z/*R, *j/*s, *j/*l, *S/*R, *S/*C/*L, *s/*l/*z, *z/*j/*d?, *z/*s/*S, *L > k, *j/*g
- b. Attested in few languages and reflected only in one or two attested languages: *j/*ñ, *l/*r, *ñ/*ŋ
- c. Common in Formosan languages: *q > ?, *q > 0

Table 3: Diagnostic phonological innovations in Formosan languages

	Sai Paz	Ata	Tha	Tao	Bab	Pap	Hoa	Tso	Kan	Saa	Bun	Pai	Puy	Ruk	Sir	Ami	Bas	Tro	Kav	PMP	
*p > ̰				+	+	+															
*S > *h																				•	
*j > ̰ ²				+				•	+	+	•			+							
*k > ?								+													
*k > ̰				+	•	•															
*-w/y > ̰			+	•	•	•	•														
*C > s	•	•		•		•	•														
*t/*C											•					•	•	•	•	•	•
*t/*s ²			•	•	•?	•	•														
*d/*s										•						•					
*j/*n (>n)																•	•	•	•	•	•
*j/*L										+							•	•	•		
*S/*s								+	•	•	•									•	
*S/*C	•						•														
*s/*l																	•	•			
*l/*R	•							•								•					
*n/*L			•					+	•		•						•	•	•		
*n/*ŋ			•	•	•	•															
*k/*g (>k)	? ?	? ?	? ?	? ?	? ?	? ?	? ?	+	•	•	•					?	?	?	?	?	

¹ In Kananavu and Saaroa *j > ̰ preceding or following *i.

² Non-finally, except in Thao, where it is also reflected finally.

- d. Attested in only a few languages but reflected in most of the latter: *r/*R, *ñ/*L
- e. Common in Austronesian: *R > ̰

Of the nine Formosan subgroups listed in (1), three are established by Blust on the basis of shared phonological innovations. They are:

- (3) a. Northwest Formosan (Blust adds *q > ?, but I take this to be a step towards *q > ̰ [cf Blust 1999a:46], which is common in Formosan languages and scarcely diagnostic):
 - i. *C > s (also attested in Taokas, Papora and Hoanya)
- b. Western Plains:
 - i. *n/*ŋ (but not reflected in Hoanya)
 - ii. *t/*s
 - iii. *-w, *-y > ̰, i.e. truncation of *-ay and *-aw
- c. East Formosan (Blust adds *q > ?, but see under Northwest Formosan above):
 - i. *t/*C (also attested in Bunun and Malayo-Polynesian)
 - ii. *j/*n
 - iii. idiosyncratic lexical changes shown in Table 2 (none of which is reflected in all five East Formosan languages)

Of these three groups, the most weakly supported is Northwest Formosan, as Blust himself points out. I will not repeat his discussion here. A glance at Table 3 shows only one possibly diagnostic change, *C > s, and this is also reflected in Taokas, Papora and Hoanya. It is really quite likely that Saisiyat and Pazeh are single-member primary subgroups of Austronesian.

The Western Plains group minus Thao was first recognised by Tsuchida (1982), and Blust adds Thao. The phonological evidence consists of the *t/*s merger, the truncation of *-ay and *-aw and the *n/*ŋ merger. The latter two innovations are a little problematic, as they were not inherited as categorical changes from Proto Western Plains: glide truncation is incomplete in Thao, and the *n/*ŋ merger has not occurred in Hoanya. In a paper dedicated to establishing the linguistic position of Thao, however, Blust (1996) shows that Thao also shares certain lexical innovations with Taokas, Favorlang-Babuza, Papora and Hoanya (and has a number of loanwords from Bunun). The Western Plains group thus seems well supported by this combination of phonological and lexical innovations.

The most clearly defined of the three groups in (3) is East Formosan, which manifests one uniquely shared innovation, the *j/*n merger (Blust 1999a:46). Blust comments that no other Austronesian language has undergone this merger, and implies that the probability is very high that it is a Proto East Formosan innovation. He writes, ‘The alternative hypothesis, that the *j/n merger in these languages is a product of two or more independent changes leaves unexplained the absence of a parallel change anywhere else in Austronesian.’

The apparent membership of Siraya in this group is, as Blust says, intriguing, located as it was on the southwestern plains, when the remaining East Formosan languages are located on the east coast. Siraya. However, it would be a somewhat wanton denial of the evidence to assume that Siraya had undergone the *j/*n and *t/*C mergers independently of the east coast languages. If one did make this assumption, then it would be entirely on grounds of geography, as the phonological evidence binding Amis into the East Formosan group is no stronger than that for including Siraya.

Blust assumes the existence of two other subgroups without providing shared innovations for them. Although there are no phonological innovations diagnostic of Atayalic, he takes the Atayalic subgroup to be self-evident (presumably because it consists of a dialect network), and I agree with him.

He also takes the Tsouic subgroup as a given, on the basis of Tsuchida (1976). Tsuchida does not list the shared innovations of Tsou, Kakanavu and Saaroa, but Table 3 displays three apparent phonological innovations linking the Tsouic languages:

- (4) a. *j > ∅: categorical in Tsou, but only preceding and following *i in Kakanavu and Saaroa; elsewhere *j > Kakanavu \l (merging with PAn *l), Saaroa \ (merging with PAn *L, *ñ) (Tsuchida 1976:306).³
- b. *S/*s merger (partial in Tsou, categorical in Kakanavu and Saaroa)
- c. *k/*g merger (partial in Tsou, categorical in Kakanavu and Saaroa)

The issues associated with the *S/*s merger are rather complex. Blust (1999a:43) shows PAn *s reflected as Tsou \s and \h, PAn *S reflected as \s only, but according to Tsuchida (1976:287, 306), PAn *s and *S both regularly become Tsou \s, except immediately preceding early Tsou \s or \t, where both become \h. If Tsuchida’s analysis is correct, then PAn *s and *S indeed merged as Proto-Tsou *s. However, in Ross (2006b), I show that the history of Tsouic sibilants was quite complex and that this merger did not take place (the relevant sound correspondences are shown in Table 6 below).

The *k/*g merger would be only weakly diagnostic, as we don’t know how many other Formosan languages reflect it (the functional load of *g was apparently so low that we lack reflexes of it in

³ PAn *j was also lost following *i in Rukaiic.

some languages). But the simplest interpretation of Tsuchida's (1976:173–176, 225-227) analysis of the fate of PAn *k and *g in Tsouic languages⁴ is that PAn *k normally becomes \ in Tsou, but in a small number of items in an earlier stage of Tsou it instead voiced to *g, becoming Tsou \k. If this interpretation is correct, then there was no *k/*g merger in Proto Tsouic.

In sum, it is possible that the Tsouic group is defined by one minor phonological innovation: PAn *j was lost adjacent to *i.

One intriguing point emerges from Table 3, namely that each of the innovations that Blust attributes to Tsouic has also occurred in neighbouring Bunun. In fact, on the face of it, the phonological argument for a subgrouping relationship between Bunun and Kanakanavu is stronger than the argument for such a relationship between Tsou and Kanakanavu. Shared innovations of Bunun and Kanakanavu are the *n/*L merger, the *k/*g merger, and the loss of *j when adjacent to *i (Bunun, like Tsou, has lost *j in all environments).⁵ This suggests that a possible relationship between Tsouic and Bunun merits further investigation.

2.2 Summing up

The listing of Formosan subgroups in (5) represents a version of the listing in (1) modified slightly to take account of the questions raised in the discussion above. However, the changes are minimal. Northwest Formosan disappears, Saisiyat and Pazeh becoming single-member groups. Tsouic is called into question, but I will have cause to discuss the unity of the Tsouic languages at several point below. And Bunun is tagged with a question for further research.

- (5) a. \Saisiyat
- b. \Pazeh
- c. Atayalic (the dialects of \Atayal and \Seediq)
- d. Western Plains (?)
 - i. Thao
 - ii. Central Western Plains (\Taokas, Favorlang-Babuza, Papora and \Hoanya)
- e. possibly Tsouic (\Tsou, Kanakanavu and \Saaroa)
- f. Rukaic
- g. \Bunun: does this form a higher-order group with Tsouic?
- h. \Paiwan
- i. \Puyuma
- j. East Formosan
 - i. \Siraya
 - ii. \Amis
 - iii. Northeast Formosan (\Ketagalan [\Basay and \Trobiawan varieties] and \Kavalan)
- k. Malayo-Polynesian

2.3 Sagart's critique of Blust's subgrouping

Sagart (2004) dismisses Blust's subgrouping on the grounds that phonological changes readily spread across language boundaries in regions where languages are contiguous with one another, giving rise to the conflicting isoglosses which, he says, are reflected in Formosan sound correspondences. If this is so, he argues, then phonological innovations shared by two or more languages are not evidence that

⁴ Tsuchida himself follows Dyen in positing PAn subscripted \k₁ and \k₂.

⁵ Table 3 also shows the *S/*s merger as common to Kanakanavu and Bunun, but in my view this merger did not take place in Kanakanavu (Ross 2006b).

they form a subgroup because the innovations did not necessarily occur in a shared ancestor but may have spread from one language to another more recently.

This dismissal is far too sweeping. Patterns of conflicting isoglosses have arisen, for example, in the German-Dutch dialect continuum because the continuum for centuries remained a single social network where the social links between component communities were only a little weaker than those within them. If we envisage an extended period early in the break-up of PAn when much of Taiwan was still united in a single dialect continuum, then, yes, it is possible, indeed likely, that conflicting isoglosses would have arisen during that period (assuming that such an extended period occurred). But it is probable that early Formosan agricultural communities were small and parochial, and that any post-PAn social network would have broken up fairly soon into small speech communities with only limited social links with each other. As the links between communities became weaker, the probability of innovations crossing community boundaries would also have lessened.

The issue of just what kinds of sound change cross what kinds of speech community boundary is one that has received relatively little theoretical attention.⁶ The existence of areal phonological features shows that changes do sometimes cross language boundaries, but such changes tend to be phonetic, like retroflexion in India or the reanalysis as tone in southeast Asia of various contrasts which have a pitch difference as a by-product. Blust himself notes that the ‘echo vowels’ of Tsouic and Rukaiic and the preglottalised consonants of Thao, Bunun and Tsouic appear to be areal phonetic features (Blust 1999a:35). The circumstances under which such spreads occur needs investigation, as there are numerous cases of contact-induced change which have given rise to radical morphosyntactic change but \not to phonological change. There is in any case scant evidence that changes in the phonological system itself, such as mergers, readily cross language boundaries.⁷

If the mergers listed by Blust spread across the early Austronesian dialect continuum, then they would probably have had to do so while social links across the community remained quite strong. It seems unlikely that this period would have lasted long enough, however, for the plethora of changes in Table 1 to have occurred, and far more likely that many of these shared innovations do reflect common ancestors, as Blust proposes.

3 Ho’s subgrouping

3.1 Review

The phonological innovations upon which Ho’s (1998) subgrouping is based are shown together with the subgrouping itself in (6). I have taken the liberty of translating the symbols which Ho uses for reconstructed PAn phonemes into the now more familiar symbols used by Blust (1999b and other publications). Ho uses Li’s (1985) PAn reconstruction, which is based in part on that of Dahl (1973) and partly on that of Tsuchida (1976), rooted in turn in Dyen (1971).

Ho discusses at some length mergers of PAn *d or *D with *z,⁸ but does not regard them as diagnostic. Blust (1999a) does not reconstruct a distinction between PAn *d and *D, attributing the reflexes of both to *d, but he also considers the *d/*z merger to be non-diagnostic.

- (6) a. Saisyat-Pazeh-Atayalic (*d/*D)

⁶ It has received much less attention than the difference between changes which are lexically categorical (affecting all lexical items in which the relevant segment occurs) and those which are lexically gradual (Labov 1981, Labov 2006).

⁷ Sagart’s one non-Formosan example is the (purely phonetic) spread of uvular /r/ across European languages. This was a city-to-city spread, initially jumping the intervening countryside because it was transmitted as a status marker (Trudgill 1983:56–59)—a situation hardly likely to have been typical of Formosan languages at any stage in their history.

⁸ PAn *z is notated by Ho (following Li 1985) as *d’/*Z (or *z), but both Ho and Li show their reflexes in Formosan languages as being identical.

- i. \Saisiyat
- ii. \Pazeh
- iii. Atayalic (the dialects of \Atayal and \Seediq)
- b. Western Plains (*t/*s, *n/*ŋ)
 - i. \Thao
 - ii. Central Western Plains (*k > 0, *w > 0)
 - A. \Taokas, Favorlang-Babuza
 - B. \Papora, Hoanya
- c. Rukai-Tsouic (*s/*S)
 - i. Tsouic (\Kanakanavu, \Tsou and \Saaroa (*D/*z, *k/*g))
 - ii. Rukaic
- d. \Paiwan
- e. \Puyuma
- f. Atypical Formosan (*t/*C)
 - i. \Siraya
 - ii. Amis-Bunun-Kavalan
 - A. \Amis
 - B. Bunun-Kavalan (*n/*ñ/*L)
 - 1. \Bunun
 - 2. \Kavalan

Ho does not mention Proto-Malayo-Polynesian, but by implication, it would fall into a group with Atypical Formosan, as it shares the *t/*C merger.

Ho recognises a Saisiyat-Pazeh-Atayalic subgroup on the basis of the *d/*D merger. The distinction between PAN *d and *D dates back in one form or another to Dahl (1973), but Blust (1999a) argues that it is an illusion brought about by borrowings from Puyuma into Paiwan. However, according to Li (1985) Saaroa, Rukaic (Budai, Maga), Sakizaya Amis, Taokas, Favorlang-Babuza and Papora also reflect the distinction, and this is taken up in Ho's discussion. In Ross (1992) I claimed that *d (my *d₁) and *D (my *d₂ and *d₃) do indeed have distinct reflexes in a number of Formosan languages, and upon re-inspection of the data I think that this claim remains valid (Ross 2006b). If I am right, then Saisiyat, Pazeh and Atayalic do indeed share a merger. However, it is also shared by Rukaic, Tsou, Saaroa, Siraya, Amis and Bunun, and so it is scarcely diagnostic as an innovation.⁹ In these circumstances, the Saisiyat-Pazeh-Atayalic subgroup lacks definitive support.¹⁰

The two mergers which Ho notes for Western Plains (*t/*s, *n/*ŋ) are among the three identified by Blust. For Central Western Plains Ho notes two innovations, *k > 0, *w > 0. Blust shows loss of *k as only partial in Taokas, and as not occurring at all in Hoanya. From the relevant data in Tsuchida (1982), listed in Table 4, we see that Taokas and Hoanya behave in the same way: both lose *k except in PAN *kita 'see', where it is retained (Pazeh is offered by way of contrast, as it retains *k). PAN *ikuR 'tail' is exceptional in that *k is apparently retained, but it is also exceptional in other respects and its history is not clear.

⁹ Li (1985) disagrees with me. On his analysis PAN initial *d- and *D- are reflected respectively as \d- and \D- in Budai and Maga Rukai, and PAN *d and *D are reflected respectively as \c and \s in Saaroa (a Tsouic language). He shows no Siraya reflex of PAN *d. He shows them merging as Bunun \d.

¹⁰ Within the subgroup, Ho separates Saisiyat from Pazeh and Atayalic. It is apparently not his intention to place Pazeh and Atayalic in a subgroup, however, as they share no innovations exclusively. Instead, he makes the point that Saisiyat has undergone innovations which Pazeh and Atayalic haven't.

Table 4: Loss of PAn *k in Central Western Plains languages and Pazeh

PAn		Taokas	Babuza	Papora	Hoanya	Pazeh
*makaʔən	eat	a-man	man	man	man	məkən
*kuSa	go	m-uxa	m-usa	ma-m-oa	usa	mu-kuza
*kuCu	louse	usu	ocho	uθu	usu	kusu
*kaRaŋ	crab	yaxan	aggan	aqan	—	kaxaŋ
*kita	see	k < um > ita	m-ita	—	k < am > ita	mi-kita
*ikuR	tail	kikul	kier	qiul	—	—
*bukeS	hair	—	—	bus	bud	bekes
*aku	I	yau	—	—	—	yaku?

Table 5: Loss of PAn *w in Central Western Plains languages

PAn		Taokas	Babuza	Papora	Hoanya
*wasu	dog	outo	atu	hato	atu
*walu	eight	—	—	halu	mi-alu
*Siwa	nine	—	—	me-siya	a-sia
*ka-wiRi	left	—	kairi	aili	—

Data relevant to Central Western Plains loss of PAn *w in Tsuchida (1982) are scant, as PAn *w had a fairly low functional load. They are listed in Table 5 (where Papora \aili ‘right’ apparently reflects PAn *ka-wiRi ‘left’). The reflexes indicate that PAn *w was lost initially in Taokas (there is no medial reflex), initially and medially in Babuza and Hoanya, and was reflected as Papora initial \h- and lost medially.

Thus the loss of neither *k nor *w is categorical. However, *k is not lost in any other Formosan language and the loss seems to have been ongoing as the Central Western Plains languages diverged. Kananavu and Saaroa have also lost *w, but Tsou has not, implying that *w had not been lost in Proto-Tsouic and that its loss in the Central Western Plains languages was therefore an independent event. Word-initially, at least *w seems to have become [h-], retained in Papora but then lost in Taokas, Babuza and Hoanya. Together, the two losses form reasonable evidence of subgrouping.

Ho’s Rukai-Tsouic grouping is defined by the merger of PAn *s (Li’s *t’) and PAn *S (Li’s *s). The situation here is complicated by the reconstruction of PAn *θ by Tsuchida (1976:127–131) (following Dyen 1971), reflected only in Rukaic and Tsouic. The relevant sibilant phonemes reconstructed by Li, Ho and Blust and their supporting sound correspondences are shown in Table 6. We see that when Ho writes of a merger of PAn *s and *S, the *s he refers to excludes reflexes attributed to PAn *θ. I interpret Tsuchida’s finding differently from him and propose that there was no PAn *θ but that PAn *s underwent a phonemic split in just Rukaic and Tsouic (Ross 2006b). If I am correct in attributing the *s/*θ distinction to a phonemic split in Proto Rukai-Tsouic (PRuTs), then Ho’s claim can be restated as a claim to the effect that those reflexes of PAn *s which reflect post-split PRuTs *s merged with PRuTs *S. On my reading of Table 6 this merger did not occur, as Kananavu reflects PRuTs *s as ∅ but *S as \s. However, the phoneme split that have rise to PRuTs *s and *θ is itself

Table 6: Reflexes of putative PAn *s, *θ and *S according to Li/Ho, Tsuchida and Blust

Li/Ho	*t'	*θ	*s	Blust	
Tsuchida	*s	*θ	*S ₁ , *S ₃	*s	*S
Tsou	s	s	s	s	s
Kanakanavu	∅	s	s	s	s
Saaroa	∅	s	∅	s	s
Proto-Rukaic	?	*θ	*s	*θ	*s
Budai Rukai	∅	θ	s		
Maga Rukai	s	θ	s		
Mantauran Rukai	∅	s	?		

further evidence for Ho's Rukai-Tsouic subgroup.

Ho does not dwell on his Atypical Formosan group, apparently because he considers it to be more distant from the rest of the Formosan languages (the 'typical' Formosan languages). It shares with Malayo-Polynesian the merger of PAn *t and *C. Insofar as it is treated as diagnostic of Malayo-Polynesian, it cannot be denied diagnostic status here.

As noted in (6), Ho's (1998:159) family tree shows its membership dividing first into Siraya on the one hand and Amis, Bunun and Kavalan on the other. There is no discussion of what defines the latter group, which in turn divides into (i) Amis and (ii) Bunun and Kavalan, the latter pair being defined by the *n/*ñ/*L merger. This is really two mergers, probably *ñ/*L, then *n/*L. The *ñ/*L merger was listed in (2d) as non-diagnostic. The merger *n/*L has greater diagnostic value, but is also found in Thao and Kanakanavu. The possibility of independent parallel innovation is thus quite strong, the more so as Bunun and Kavalan are distant from each other. On balance, it seems improbable that they form a subgroup.

3.2 Summing up

The grouping in (7) represents a version of Ho's subgrouping, modified in the light of the discussion above. The Saisiyat-Pazeh-Atayalic subgroup disappears, as does the internal subgrouping within Atypical Formosan.

- (7)
- a. \Saisiyat
 - b. \Pazeh
 - c. Atayalic (the dialects of \Atayal and \Seediq)
 - d. Western Plains
 - i. \Thao
 - ii. Central Western Plains
 - A. \Taokas, Favorlang-Babuza
 - B. \Papura, Hoanya
 - e. Rukai-Tsouic
 - i. Tsouic (\Kanakanavu, \Tsou and \Saaroa)
 - ii. Rukaic
 - f. \Paiwan
 - g. \Puyuma
 - h. Atypical Formosan

- i. \Bunun
- ii. \Siraya
- iii. \Amis
- iv. \Kavalan

Significantly, this revised subgrouping is similar to the one in §2.2 which resulted from reviewing Blust’s work. The differences are (i) that Ho recognises a Rukai-Tsouic subgroup where Blust has separate Tsouic and Rukaic groups; (ii) where Blust’s East Formosan includes Kavalan, Siraya and Amis, Ho’s Atypical Formosan also includes Bunun (which, I suggested in §2.1 shows some affinities with Kanakanavu).

4 Sagart’s subgrouping

4.1 Review

Whereas Blust’s subgrouping consists of nine apparently primary subgroups of Austronesian in Taiwan and Ho’s has six (ten and eight respectively with my tentative modifications), Sagart’s (2004), shown in (8) has only three, of which one, Pituic, includes most of the languages of the family and undergoes recursive splits. It includes Tai-Kadai (alias Kra-Dai), incorporating the claim that Tai-Kadai languages form a subgroup of Austronesian. This is an important claim, but one that I will not address here.

- (8) a. \Saisiyat
- b. \Pazeh
- c. **Pituic**
 - i. Atayalic (the dialects of \Atayal and \Seediq)
 - ii. \Thao
 - iii. \Taokas
 - iv. \Favorlang-Babuza
 - v. \Papora
 - vi. \Hoanya
 - vii. **Enemic**
 - A. \Siraya
 - B. **Walu-Siwaic**
 - 1. Rukai-Tsouic
 - Rukaic
 - Tsouic (\Tsou, Kanakanavu and \Saaroa)
 - 2. \Bunun
 - 3. \Paiwan
 - 4. \Puyuma
 - 5. \Amis
 - 6. **Muic**
 - Northeast Formosan (\Ketagalan and \Kavalan)
 - Tai-Kadai
 - Malayo-Polynesian

Sagart accepts three low-level groups without further discussion: Atayalic as self-evident, Rukai-Tsouic on the basis of lexical innovations presented by Tsuchida (1976:11–12), and Northeast Formosan on the basis of innovations presented by Blust (1999a) and Li (2001). Atayalic is indeed self-evident, and so is Northeast Formosan.

Sagart founds his larger subgrouping argument on the fact that the conventionally reconstructed PAN numeral forms from ‘5’ to ‘9’ are not reflected in certain Formosan languages (see Table 7). He argues that *pitu ‘7’, *walu ‘8’ and *Siwa ‘9’ did not occur in PAN. Instead, there were only multiplicative (‘twice three’, ‘twice four’), additive (5 + 1 etc), and—in the case of ‘nine’—subtractive (10 – 1) numerals, and, he argues, *pitu, *walu and *Siwa were post-PAN abbreviated forms of PAN additive numerals. These abbreviations constitute innovations at post-PAN interstages and are the main planks of Sagart’s subgrouping: hence the subgroup labels ‘Enemic’ and ‘Walu-Siwaic’.

Sagart’s (2004:414) data are given in Table 7. They are supplemented here by reconstructions and some additional conservative dialect forms from Ferrell (1969) and, for Taai Saisiyat, from Li (1978). One significant difference between my presentation and Sagart’s is that I have placed Hoanya and Papora under Pituic rather than under the lower-order Walu-Siwaic, firstly because Sagart himself excludes them from the latter group and secondly in anticipation of my decision to subgroup them with Favorlang and Taokas in a Central Western Plains group. Another difference is that I have shifted Kavalan and Ketagalan out of the group of languages reflecting *puluq ‘ten’, since they don’t reflect it.

Passing over certain details, Sagart’s argument is as follows. He reconstructs PAN *RaCeb¹¹ ‘five’ on the basis of Pazeh \xasep, Saisiyat \aaseb, Favorlang \achab, Taokas \hasap. Favorlang \achab and Taokas \hasap appear to be borrowings. The expected forms are respectively *\rachab and *\lasab. *RaCeb is thus directly reflected only in Pazeh and Saisiyat. Sagart then extrapolates from the Pazeh additive numerals for 6–9 to reconstruct the (probably) PAN additive forms shown in (9). He then proposes that (i) *RaCeb-i-Sepat became *RaCeb-i-Sipat (vowel harmony); (ii) schwa (orthographic *e) was deleted; (iii) there was an alternant form of *duSa ‘2’, namely *tuSa (Amis \tusa, Rikavong Puyuma \toa¹²); (iv) final *-a of *duSa/*tuSa was originally the ligature, giving penultimate stress where other numerals had final stress. These proposals lead to the ‘Stage X’ reconstructions in (9). These forms then underwent pruning to the left of the pretonic syllable and to the right of the stressed vowel, as well as lenition of *-pa- to *-wa- and reduction of *-tl- to *l to give the ‘Stage Y’ reconstructions in (9). The parts of the Stage X reconstructions reflected in Stage Y are shown in bold. The series of changes that Sagart posits here is perhaps plausible but, as he admits, idiosyncratic.

(9)

	‘6’	‘7’	‘8’	‘9’
PAN ?	*RaCeb-esa	*RaCeb-i-duSa	*RaCeb-a-telu	*RaCeb-i-Sepat
Pazeh	<i>xasebuza</i>	<i>xasebidusa</i>	<i>xaseb[a,i]turu</i>	<i>xasebisupat</i>
Sekhwan Pazeh	<i>buda</i>	<i>biduset</i>	<i>bituro</i>	<i>bisupat</i>
Luilang	<i>patulu-nai</i>	...
Makatao Siraya	<i>sipat</i>
Stage X		*RaCb-i-túS(a)	*RaCb-a-tlú	*RaCb-i-Sipát
Stage Y		*pitu	*walu	*Siwa

¹¹ Sagart reconstructs *RaCep, a difference in detail that is not relevant here.

¹² Sagart adds Thao \tu a, but this is a regular reflex of PAN *DuSa ‘2’: Ross 2006b.

Table 7: Numeral forms in Formosan languages and PMP

	‘7’	‘5’	‘6’	‘8’	‘9’	‘10’
PAAn ?	*RaCeb- i-duSa	*RaCeb	*RaCeb-esa *[ma][kaR]tolu	*RaCeb-a-telu *[ma][kaR]Sepat	*RaCeb-i-Sepat *(t,s)a(n,ŋ)aCu	*[sa]iCit
Saisiyat	saivuseaha	rasu	saivusa	makaspat	ra:ha	ranpon
Taai	ʃayboʃiro ʔəhæʔ	rasəb	ʃayboʃir	kaʃpat	ræ:ʔhæʔ	laŋpəz
Pazeh	xasəbidusa	xasəp	xasəbuza	xasəb[a,i]turu	xasəbisupat	isit
Proto-Pituic	*pitu	? *lima	? *enem	? *walu	? *Siwa	
Favorlang	naito	achab	nataap	maaspat	tannacho	zchiett
Taokas	yuweto	hasap	tahap matu	mahalpat, makaipat	tanasu	ta-isid
Hoanya	pito	ʎima	mi-num	mi-alu	a-sia	miata-isi
Papora	pitu	nema	[me]nom	mahal, halu	[me]siya	[me]tsi
Thao	pitu	*lima rima	katuru, makaʔturuturu	kaʃpat, maka(ʔ)ʃpaʃpat	tanaθu	*masehaL maqθin
Atayal	pituʔ	ima-gal	cziuʔ	[ma]spat	qeru	lpuu
Seediq	pito	lima	mataro	maspat	maŋali	mahal
Proto-Enemic			*enem			
Siraya	pʰittu	rima	nəm	kuiɣpa, kuyipat	matuda	saat kʰittian
Proto-Walu-Siwaic				*walu	siwa	
Tsou	pítu	eímo	nómə	vóeu	sífo	máskə
Saaroa	[ku]pito	[ku]lima	[k]ənəmə	[k]ualo	[ku]sia	[k]uma:ʔə
Kanavu	pitu	rima	nəm	(h)a:ru	si:ya	ma:nə
Rukai	pitu	ʎima	eneme	vaʎu	baŋatə	maeale
Bunun	pituʔ	himaʔ	nuum	vauʔ	sivaʔ	masʔan
Kavalan	pitu	rima	nem	waru	siwa	betin
Ketagalan	pitu	tsjima	anəm	wasu	siwa	labatan
Paiwan	pitju	lima	[u,e]nem	alu	siva	*puluq puluq
Puyuma	pitu	ʎima	nem	waʎu	iwa	puʎu
Amis	pitu	lima	enem	falu	siwa	polo
PMP	*pitu	*lima	*enem	*walu	*siwa	*puluq

We would expect the proposed intermediate stages to be reflected at least fragmentarily in some Formosan languages, and Sagart cites the Sekhwan Pazeh in (9) as left-truncated versions of the Pazeh forms.¹³ The Sekhwan Pazeh forms do not fit regularly into Sagart's history but do illustrate the possibility of left-truncation. Old Pazeh \biduset '7' (for expected *bidusa) is puzzling, and \bisupat (for expected *s(u,i)pat) has an unexpected initial syllable.¹⁴ The expected \sipat is recorded for Makatao Siraya, however. Luilang was apparently a Western Plains language, but just a few lexical items survive, including \in-nai '7', \patulu-nai '8' and \satulu-nai '9'. Sagart treats \patulu- as reflecting left-truncation of *RaCeb-a-telu. He may be right, but our ignorance of Luilang historical phonology and the failure of '7' and '9' to reflect truncated forms leave room for skepticism.

The form *pitu '7' is reflected in all Formosan languages except Saisiyat and Pazeh, but *walu '8' and *Siwa '9' are reflected in a more constrained set which also excludes Atayalic, Taokas, Favorlang, Thao and Siraya. Sagart argues that *pitu emerged earlier as the default form for '7' because there were no multiplicative or subtractive alternatives for it, whereas *walu and *Siwa won out over the latter only later. Thus in Table 7 Taokas \maha.lpat, Favorlang \maaspat, Thao \ka pat and \maka() pa pat, Atayal \spat and Seediq \maspat all reflect PAn (?) *[ma][kaR]-Sepat 'eight', consisting of PAn *Sepat '4' prefixed by a probable multiplicative prefix or prefixes.¹⁵ Taokas \tanasu, Favorlang \tannacho, Thao \tanacu reflect a protoform *(t,s)a(n,t)aCu 'nine' which Sagart takes to be possibly subtractive.

Sagart labels the language in which *pitu '7' first appeared 'Pituish' and the (lower-order) language in which *walu '8' and *Siwa '9' emerged 'Walu-Siwaish'. His naming convention is to use labels in \-ish for interstage languages and labels in \-ic (Pituic etc) for subgroups. In what follows I replace the former with labels in the more conventional format of 'Proto-Pituic' etc, as this makes the distinction between protolanguage and subgroup more salient.

Although Papora and Hoanya have reflexes of *walu and *Siwa,¹⁶ Sagart attributes these two languages only to Pituic, not to Walu-Siwaic, on the grounds that their reflexes of *walu and *Siwa reflect PAn *w as zero (Papora \maha.l and \me-siya, Hoanya \mi-alu and \a-sia and are therefore probable borrowings from a Tsouic language. This presupposes that PAn *w was otherwise retained in Papora and Hoanya, a proposal for which Sagart offers no evidence. Table 5 shows that PAn *w was in fact lost in Central Western Plains languages except initially in Papora, where it was retained as \h-, suggesting that Papora \maha.l, \halu and Hoanya \mi-alu 'eight' are directly inherited reflexes of PAn *walu, and Papora \[me]sia and Hoanya \a-sia of PAn *Siwa. Papora and Hoanya thus meet the criterion for membership of Walu-Siwaic. This seems problematic, as their closest relatives, phonologically and lexically, are Favorlang-Babuiza and Taokas, which do not meet this criterion.

The difficulty is perhaps not as large as it looks, as there is a weakness in Sagart's argument at this point. The attrition of the Stage X forms in (9) to Stage Y is attributed to a single set of processes which would presumably have affected the three numerals simultaneously, yet Sagart invites us to believe that one form, *pitu, won out before the other two, *walu and *Siwa. If I understand Sagart correctly, what this means is that *walu and *Siwa or their immediate precursors must also have occurred in Proto-Pituic (i.e. Stage Y was Proto-Pituic), but alongside their multiplicative or

¹³ The Sekhwan Pazeh forms are cited by Sagart from Imbault-Huart (1893:319), who cites them from Thomson (1873). The same forms are cited in Li and Toyoshima (2006), where the location is noted as Sekhwan. I have rendered them into a more modern orthography for the sake of comparability.

¹⁴ *Raceb '5' is reflected as Sekhwan \hasub (Li and Toyoshima 2006:649).

¹⁵ One or both of the square-bracketed prefixes is reflected in each reflex except Atayal.

¹⁶ The fact that *Siwa is irregularly reflected as PMP *siwa is discussed by Blust (1995b) and is irrelevant to the issues at stake here.

subtractive counterparts, which were banished only in Proto-Walu-Siwaic. I note this hypothesis with question-marked Proto-Pituic protoforms in Table 7. The weakness in the argument lies in the fact that there is a high probability of the independent loss/retention of one or the other of a pair of alternative lexical items. That is, neither the absence of reflexes of *walu and *Siwa in certain Pituic languages (Atayalic, Taokas, Favorlang, Thao, Siraya) nor their presence in others is necessarily significant for subgrouping, and the boundaries of a separate Walu-Siwaic group are thus not well defined.

Much the same is true of the replacement of *RaCeb ‘5’ by reflexes of *lima ‘hand’, and of various forms for ‘6’ by reflexes of *enem, an etymon of unknown origin. Sagart points out that it is quite possible that *lima and *enem were already present in Proto-Pituic, although neither is reflected in the available data for Taokas and Favorlang, and *enem is also missing from Atayalic and Thao. He opts for positing Enemic as a subgroup, but not ‘Limaic’. The reason for the difference between these two decisions is not clear to me, but Enemic is also supported by the replacement of PAn *kawaS ‘year’ by reflexes of *CawiL.

For ‘10’, Formosan languages display three forms with multiple reflexes:

- (10) a. PAn *[sa]jCit:¹⁷ Pazeh \isit, Taokas [\ta] isid, Favorlang \zehiett, Papora \[me] tsi, Hoanya \[miata] isi
 b. *masehaL (Tsuchida 1976: *mas?aL):¹⁸ Seediq \mahaL, Tsou \másk, Saaroa \ma: , Kakanavu \ma:n, Bunun \mas an, Rukai \maeale
 c. *puluq: Paiwan \puluq, Puyuma \pu u, Amis \poLo, PMP *puluq

At one extreme, Pazeh, external to Pituic, reflects PAn *[sa]jCit.¹⁹ At the opposite extreme *puluq only occurs in a subset of Walu-Siwaic. However, between these extremes, as Sagart indicates, there must have been alternations among the three etyma. Walu-Siwaic languages reflect all three etyma, and Pituic languages external to Walu-Siwaic reflect the first two. This highlights the large risk that Sagart takes in using numeral forms as a basis for subgrouping.

Within the Walu-Siwaic subgroup, Sagart recognises a Muic subgroup, so labelled because its members are said to reflect that part of Blust’s ‘second politeness shift’ whereby the 2PL genitive enclitic PAn * = mu came to be used as 2SG genitive. Blust (1977) claims this as a defining innovation of Malayo-Polynesian, but Sagart promotes it to a defining innovation of a subgroup that also includes Northeast Formosan and Tai-Kadai. I will not consider the latter here. Of immediate concern is the claim that the Northeast Formosan languages Ketagalan and Kavalan are Muic. The basis of Sagart’s claim is a sentence in Li (1995:667):

Moreover, Trobiawan, a variety of Ketagalan, uses the form \imu ‘your (sg.)’ as in \tama-imu ‘your father’ (Asai [], Text 6, rather than \ (i) su as in most other Formosan languages.

The text to which Li refers does indeed contain \-imu as a 2SG possessor pronoun, but after examining further texts Li (1996:179, 1999:639, 643, 663) concludes that in the Basay variety of Ketagalan \[i]mu was the 2PL form and \[i]su the 2SG. That is, the politeness shift had not occurred in Basay. We know that it has not occurred in Kavalan, and it therefore did not occur in Proto-Northeast Formosan. What Li observed in one Ketagalan text does not seem to reflect the politeness shift in a putative Proto-Muic but rather, perhaps, an occasional polite use of a plural for a singular. The second politeness shift reconstructable for Proto-Malayo-Polynesian was in fact a rather complex set of changes (Ross 2006a), and a Muic subgroup would need to reflect the set, rather than a single plural-

¹⁷ PAn *sa- ‘one’.

¹⁸ The reconstruction is problematic, as Sagart mentions, because some sound correspondences are odd. Possibly Thao \maqcin is also cognate.

¹⁹ Saisiyat, also external to Pituic, has \ranpon, of unknown origin.

to-singular shift. The latter is such a common strategy as to have little value as subgrouping evidence (Blust 1995a:621).

Sagart accepts Tsuchida's Rukai-Tsouic group, which is based on the (PRuTs) lexical replacements listed in (11), reconstructed by Tsuchida (1976:11–12).

(11)	*tukuLu	'heart'
	*DakəraLə	'river'
	*CaŋəRaLə	'star'
	*-bali	'smell'
	*ramuCu	'hand'
	*S ₁₆ iqipi	'shoulder'
	*qaputu	'hammer'
	*k(u,ə)k(u,ə)	'leg'
	*(n,t)ətənə	'lungs'
	*ŋuRuq ₂ u	'nose' or 'nasal mucus'

The last item on the list, putative Proto-Rukai-Tsouic *ŋuRuq₂u, appears to have cognates in the Northwest Solomonian languages: Babatana, Roviana and Nduke \ŋuru 'nasal mucus'. If so, then it is not a Rukai-Tsouic innovation. Tsuchida also reconstructs *La 'and' and *Si 'because', but I have omitted these from (11) because the borrowing of conjunctions is very common in bilingual societies. This leaves nine putative shared lexical replacements in (11). The in-principle difficulty with lexical replacements is that one can never be quite sure that cognates will not be uncovered elsewhere, as in the case of *ŋuRuq₂u. However, five of these replacements are body-parts, and 'star' is also a piece of basic vocabulary. It is distinctly unlikely that there are still cognates waiting to be found. Another possibility is that these are borrowings, but that is unlikely as these are basic vocabulary items. Provisionally, therefore, I accept the hypothesis that Rukai-Tsouic is a subgroup, but one that is not as well supported as it would be if there were phonological, morphological and/or idiosyncratic lexical innovations

4.2 Summing up

I have expressed a certain skepticism above about the evidence for the Enemic and Walu-Siwaic subgroups, together with a view that there is no evidence for a Muic subgroup of Austronesian. This does not mean that I reject Sagart's findings out of hand, but that I am less certain than he is about their interpretation. It is clear that neither Saisiyat nor Pazeh undergoes any of the innovations that he discusses,²⁰ and conversely that all other Formosan languages reflect the innovation that gave rise to *pitu '7'. It is very likely that the directionality of the innovations in numerals is indeed as he reconstructs it (but this does not mean that the innovatory processes Sagart reconstructs are necessarily correct). It is also apparent that Rukaic, Tsouic, Bunun, Paiwan, Puyuma, Amis, Northeast Formosan and Malayo-Polynesian have undergone the innovations which gave rise to *lima '5', *enem '6', *walu '8' and *Siwa '9'. Only Paiwan, Puyuma, Amis and Malayo-Polynesian reflect *puluq '10', but Sagart does not argue that this has subgrouping significance.

The difficult area, as I observed above, is that occupied by Atayalic, Thao, Taokas, Favorlang-Babuza, Papura, Hoanya and Siraya, where it is necessary to posit pairs of alternating lexical items, with the survival probability of each being about equal. At this probability level, independent parallel

²⁰ Nor apparently does Luilang, but we have so little data for this language that I will not discuss it.

innovation is so likely that the innovations have no value for subgrouping purposes. Thus the evidence offered for Enemic and Walu-Siwaic is rather weak.

In recent work on the history of Formosan case-markers and pronouns (Ross 2006a), I found certain forms that are reconstructable to Proto-Pituic and to Proto-Enemic, but none reconstructable to Proto-Walu-Siwaic or Proto-Muic.

The apparent Proto-Pituic innovations are:

- (12) a. *Ci ‘proper-name marker’: Thao *θi-* (in *θiθu* ‘free 3s pronoun’, Siraya *ti*, Paiwan *ti*,²¹ Bunun *-t* (in *[k]a-t* ‘personal nominative case-marker’), Kavalan *ti*
- b. *i ‘common locative case-marker’: P-Atayal **i*?, Favorlang *i*, P-Puyuma **i*, P-Amis **i*, P-Bunun **i* ‘oblique case-marker’, PMP **i* LOC
- c. *-an ‘locative case-marker’: P-Atayal **-an* ‘personal oblique case-marker’, Siraya *-an* ‘personal oblique case-marker’, P-Amis **-an* ‘locative case-marker, personal accusative case-marker’, P-Rukai **-anə* ‘personal oblique case-marker’, Kavalan *-an* ‘locative case-marker’
- d. the nominative pronominal enclitics. This may mean that the encliticisation of nominative forms had not occurred in PAn and occurred only after the ancestor of the Pituic subgroup had separated from the ancestors of Pazeh or Saisiyat.

A weakness of all the above as subgrouping evidence is that they may have occurred in PAn and simply been lost in Pazeh and Saisiyat.

The apparent Proto-Enemic innovations are:

- (13) a. the genitive personal plural case-marker **ni-a*, reflected by Paiwan *nia* and Amis *na* (contrasting in both languages with singular *ni*);
- b. polite first- and second-person pronoun forms in **k-*, reflected in Siraya, Kanankanavu and Malayo-Polynesian.

In neither case are reflexes found right across Enemic, and so these cannot be regarded as strong subgrouping evidence.

The outcome of this discussion is that Sagart’s findings lead me to a subgrouping tree which is somewhat flatter than his, as shown in (14).

- (14) a. \Saisiyat
 b. \Pazeh
 c. **Pituic ?**
 i. Atayalic (the dialects of \Atayal and \Seediq)
 ii. \Thao
 iii. \Taokas
 iv. \Favorlang-Babuza
 v. \Papora
 vi. \Hoanya
 vii. **Enemic ??**

²¹ The Paiwan form is *ti* instead of expected **tsi*, and this probably reflects a formal conflation with Paiwan *ti* ‘personal nominative case-marker’. These are clearly homophonous but distinct morphemes, as Paiwan *ti* ‘proper-name marker’ does not mark case but precedes a personal noun that is head of a noun phrase containing one or more attributive items. Such a noun phrase is marked with a *common* case-marker (data provided by Anna Hsiou-chuan Chang).

- A. \Siraya
- B. Rukai-Tsouic
 - 1. Rukaic
 - 2. Tsouic (\Tsou, Kanakanavu and \Saaroa)
- C. \Bunun
- D. \Paiwan
- E. \Puyuma
- F. \Amis
- G. Northeast Formosan (\Ketagalan and \Kavalan)
- H. Malayo-Polynesian

This subgrouping really resembles the flattened versions of Blust's and Ho's subgroupings in only one respect: the attribution of Saisiyat and Pazeh to single-member primary subgroups. Pituic corresponds to nothing in the other two subgroupings. Enemy is a much extended version of Ho's Atypical Formosan, with the addition of Rukai-Tsouic, Paiwan and Puyuma. On the other hand, apart from Rukai-Tsouic, shared with Ho's grouping, this subgrouping lacks smaller groups: there is no Western Plains, no East Formosan.

5 Reconciling the subgroupings

I have subjected each of the three subgroupings above to a critical review, producing a modified version of each in which ill-supported groups are eliminated. In each case, the result is a flatter family tree, the flattening being least in the case of Blust's tree, greatest in the case of Sagart's. The question to be addressed here is whether the three modified subgroupings can be reconciled. Each was produced using the classical comparative method, and the question is therefore whether the innovations posited in the modified subgroupings contradict or, as we would expect from applications of the comparative method, support one another. From this point I will simply refer to them as 'the subgroupings', referring to the modified versions in (5), (7) and (14).

I showed in §3.2 that the flattened versions of Blust's and Ho's subgroupings are quite similar. Curiously, the flattened version of Sagart's subgrouping is complementary to Blust's and Ho's. In many respects, then, the three subgroupings can be reconciled, and the outcome is shown in (15), where the parenthesised names point to the flattened subgroupings in §2.2, §?? and §4.2.

- (15) a. \Saisiyat (Blust, Ho, Sagart)
- b. \Pazeh (Blust, Ho, Sagart)
- c. Pituic ? (Sagart)
 - i. Atayalic (the dialects of \Ataya1 and \Seediq) (Blust, Ho, Sagart)
 - ii. Western Plains (Blust, Ho)
 - A. Thao
 - B. Central Western Plains (\Taokas, Favorlang-Babuza, Papora and \Hoanya) (Blust, Ho)
 - iii. Enemy ?? (Sagart)
 - A. \Paiwan (Blust, Ho, Sagart)
 - B. \Puyuma (Blust, Ho, Sagart)
 - C. Rukai-Tsouic (Ho, Sagart)
 - 1. Rukaic (Blust, Ho, Sagart)
 - 2. Tsouic (\Tsou, Kanakanavu and \Saaroa) (Blust, Ho, Sagart)

- D. \Bunun (Does this form a higher-order group with Rukai-Tsouic? Or does it belong with East Formosan?)
- E. East Formosan (Blust)
 - 1. Northeast Formosan (\Ketagalan [\Basay and \Trobiawan varieties] and \Kavalan) (Blust,Sagart)
 - 2. \Siraya
 - 3. \Amis
- F. Malayo-Polynesian (Blust, Sagart)

The three subgroupings agree that Pazeh and Saisiyat are each single-member primary Austronesian subgroups.

I have included Sagart's Pituic in (15) with a question-mark, but it is directly supported only by the lexical innovation *pitu 'seven' and the morphological innovations that I added in (12), and these are not conclusive evidence. I retain Pituic because I think it is a hypothesis worthy of further research.

Pituic falls into three subgroups: Atayalic, Western Plains and Enemic. If one rejects Pituic, then these become primary subgroups of Austronesian. Atayalic is accepted by all scholars. Western Plains, with its internal division into Thao and Central Western Plains, is recognised by both Blust and Ho. The innovation sets that they propose largely overlap. Sagart has nothing to say on this topic, the more so as on the basis of their numerals the Central Western Plains fall into two pairs, (i) Taokas and Favorlang-Babuza and (ii) Papora and Hoanya. I suggested in §4.1 that this reflects the fact that there were alternant numeral forms in early Pituic.

The third of the three divisions of putative Pituic, Sagart's Enemic, is again retained because of its value as a hypothesis. At present it is supported by just two lexical innovations, *enem 'six' and *CawiL 'year', and by the gradual accretion of innovations (*walu '8', *Siwa '9', *puluq '10') as one moves down the list of languages in (15). Because these innovations consist of one lexical alternant winning out over another, their subgrouping significance is weak and it is very difficult to know where and whether subgroup boundaries based on numerals should be drawn.

Enemic includes six subgroups: Paiwan, Puyuma, Rukai-Tsouic, Bunun, East Formosan and Malayo-Polynesian. If one rejects Enemic, then these become subgroups of Pituic, and if one rejects Pituic as well, then they become primary subgroups of Austronesian. It is about some of these subgroups that the scholars whose work is reviewed above display significant disagreements.

There is full agreement that Paiwan and Puyuma are separate single-member subgroups.

Ho posits Rukai-Tsouic on the basis of innovations in sibilants. Although I interpret the data differently from him, the phonemic split in PAn *s does support Rukai-Tsouic. Sagart, meanwhile, accepts it on the basis of lexical innovations put forward by Tsuchida. Blust does not posit a Rukai-Tsouic subgroup. If I understand him correctly, the reason for this is that he does not see the patterning of sibilant reflexes displayed in Table 6 and hence sees no shared innovation.

An additional question here concerns Bunun, which, as I noted in §2, shows innovations in common with Tsouic languages, a fact discussed in none of the subgroupings reviewed above. It does not, however, share in the sibilant behaviour that unites Rukaic and Tsouic. Rukaic, on the other hand, does not reflect the innovations common to Tsouic and Bunun. There are perhaps signs here that we are confronted with an innovation-linked rather than an innovation-defined subgroup (Pawley and Ross 1993), i.e. a grouping in which some innovations have spread after its break-up.

It is well accepted that Kavalan and Ketagalan form the small Northeast Formosan subgroup. Blust places Northeast Formosan, Amis and Siraya in an East Formosan group, supported principally by the *j/*n merger and less strongly by the *t/*C merger. Ho, on the other hand, includes Bunun with these languages in an Atypical Formosan group on the grounds that they share the *t/*C merger—and this group would presumably also include PMP.

The only significant point of disagreement, then, is over Bunun. The evidence in Table 3 places it with Tsouic, except for the */*C merger, which places it with the East Formosan languages. There is no immediately available way of resolving this dilemma.

It is, I hope, obvious that this subgrouping has not been arrived at by ‘picking and choosing’ from the three subgroupings. All three result from applications of the comparative method, and if the method is valid and the applications are competent, then their reconciliation should be possible. In §2.3 I rejected Sagart’s dismissal of Blust’s phonological innovations. What I have attempted to do here is to evaluate each of the shared innovations claimed under each subgrouping, and to assess each subgrouping without reference to the others. Interestingly, at the end of this process Blust’s and Ho’s subgroupings are quite similar, and Sagart’s emerges as largely complementary to theirs. The conflicts can be easily enumerated:

- (16) a. the position of Bunun;
- b. the integrity of Rukai-Tsouic;
- c. the existence of Pituic and Enemic;

A great deal more information in the form of shared innovations will be required before these issues can be properly resolved—if they can be resolved at all. They include the vexed question of whether certain Formosan languages are more intimately related than others to Proto-Malayo-Polynesian. The places to look for relevant shared innovations are in bound morphology and in careful lexical studies like Blust (1996).

This paper has brought together the sets of shared innovations used to define subgroups by three different scholars, has assessed them, and has examined what they contribute to the subgrouping of Formosan languages. The result does not lead to eureka moments. Instead, the subgrouping in (15) is striking for its flatness, the more so if the reader is skeptical about Pituic and Enemic.

Sagart attempts to provide us with a tree that bifurcates recursively in one direction, reflecting the geography of settlement. Roughly speaking, this is the sort of pattern which emerges in the family tree of the Oceanic subgroup of Austronesian (see Ross et al. 1998:6 or Ross et al. 2003:7). However, such a tree can be coaxed to grow out of the Formosan data only with great reluctance, reflecting the probable fact that Proto Austronesian broke up rather slowly as its speakers spread across the island. Their speech diversified into dialects and languages, but all the time remaining in some degree of contact with their neighbours (more so, of course, on the plains and on the east coast than in the mountains). This would have given rise at various places and times to innovation-linked groups like the one mentioned above, and the result would not be a neatly bifurcating tree. The picture in (15) does not offer much that is new, but it does have the kind of shape that we would expect in such circumstances.

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.
- , 1995a. The position of the Formosan languages: Method and theory in Austronesian comparative linguistics. In Li 1995, 585–650.
- , 1995b. Sibilant assimilation in Formosan languages and the Proto-Austronesian word for ‘nine’ : a discourse on method. *Oceanic Linguistics* 34:443–453.
- , 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35:272–294.
- , 1999a. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38:322–365.

- , 1999b. 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, Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica* 1, 31–94. Taipei: Academia Sinica.
- Dahl, Otto Christian, 1973. *Proto-Austronesian*. Oslo: Studentlitteratur.
- Dyen, Isidore, 1971. The Austronesian languages and Proto-Austronesian. In T. E. Sebeok, ed., *Current Trends in Linguistics, vol.8: Linguistics in Oceania*, 5–54. The Hague: Mouton.
- Ferrell, Raleigh J., 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Nankang, Taipei: Institute of Ethnology, Academia Sinica. (*Monograph* 17).
- Ho, Dah-an (Hé Dàān), 1998. Táiwǎn nándǎoyǔ de yǔyán guānxi [Genetic relationships among the Formosan languages]. *Chinese Studies* 16:141–171.
- Imbault-Huart, Camille, 1893. *L'île Formose*. Paris: Ernest Leroux.
- Labov, William, 1981. Resolving the Neogrammarian controversy. *Language* 57:267–308.
- , 2006. Transmission and diffusion. Unpublished manuscript, revised, University of Pennsylvania.
- Li, Paul Jen-kuei, 1978. A comparative vocabulary of Saisiyat dialects. *BIHP* 49:133–199.
- , 1985. The position of Atayal in the Austronesian family. In Andrew K. Pawley and Lois Carrington, eds, *Austronesian linguistics at the 15th Pacific Science Congress, 257–280*. Canberra: Pacific Linguistics.
- , 1995. Formosan vs. non-Formosan features in some Austronesian languages of Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang and Cheng-hwa Tsang, eds, *Austronesian studies relating to Taiwan, Symposium Series of the Institute of History and Philology, Academia Sinica* 4, 651–681. Taipei: Institute of History and Philology, Academia Sinica.
- , 1996. *Yílán Xiàn Nándǎo mínzú yǔ yǔyán [The Austronesian ethnic groups and languages of Yilan County]*. Ilan: Yílán Xiàn Zhèngfǔ [Ilan County Government].
- , 1999. Some problems in the Basay language. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds, *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica. (*Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica* 1).
- , 2001. Bāsài yǔ de dìwèi [The position of the Basay language]. *Language and Linguistics* 2:155–171.
- Li, Paul Jen-kuei and Masayuki Toyoshima, eds, 2006. *A comparative vocabulary of Formosan languages and dialects by Naoyoshi Ogawa*. Tokyo: Research Institute of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Pawley, Andrew K. and Malcolm D. Ross, 1993. Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22:425–459.
- Ross, Malcolm, 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31:23–64.
- , 2006a. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Henry Y. Chang, Lillian M. Huang and Dah-an Ho, eds, *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on His 70th Birthday*, 521–564. Taipei: Institute of Linguistics, Academia Sinica. (*Language and Linguistics Monograph Series* W5).
- , 2006b. Some Proto Austronesian coronals reexamined. Unpublished ms, The Australian National University.

- Ross, Malcolm, Andrew Pawley and Meredith Osmond, eds, 1998. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society, Pacific Linguistics 1*. Canberra: Pacific Linguistics.
- , 2003. *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society. 2: The physical world*. Canberra: Pacific Linguistics.
- Sagart, Laurent, 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43:411–444.
- Thomson, J., 1873. Notes of a journey in southern Formosa. *Journal of the Royal Geographical Society* 43:97–107.
- Trudgill, Peter, 1983. *On dialect: Social and geographical perspectives*. New York: New York University Press.
- Tsuchida, Shigeru, 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa. (*Studies of Languages and Cultures of Asia and Africa, Monograph Series* 5).
- , 1982. *A Comparative Vocabulary of Austronesian Languages of sinicized ethnic groups in Taiwan, Part I: West Taiwan*. Tokyo: Faculty of Letters, University of Tokyo. (*Memoirs of the Faculty of Letters, University of Tokyo* No.7).