

SUBGROUPS IN AUSTRONESIAN

Mark Harvey

0. NOTE

This is a considerably revised version of the paper presented to the Third International Conference on Austronesian Linguistics. The elaboration of some points owes much to the comments and criticism of such people as Robert Blust, Lawrence Reid, Raleigh Ferrell and David Zorc, to whom must go my gratitude. Especial thanks are due to William Foley for his valued remarks. I hasten to add, of course, that nonetheless any faults remain my own.

1. THE PAN PHONOLOGICAL SYSTEM

The PAN¹ phonological system has been and continues to be the subject of dispute in certain of its areas. The vowel system and a large part of the consonant system have been reconstructed with a reasonable degree of certainty. The phonemes which are generally accepted are:

a) Vowels	i	u				
	ə					
	a					
b) Consonants						
	p	t	ts	t'	k	q
	b	d	dz			
		s			ɣ	h
	m	n			ŋ	
		l, ɭ				

ts = t₂ (Dahl) C(Dyen)

dz = d₂ (Dahl)

t' = s (Dyen)

s = S (Dyen)

However, when one attempts to reconcile all the apical and palatal reflexes in the various languages, immense problems arise. This is the area where the

Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, vol.2: *Tracking the travellers*, 47-99. *Pacific Linguistics*, C-75, 1982.

© Mark Harvey

reconstructions vary greatly and where the dispute has been strongest. If one follows the usual principles of historical linguistics by positing a proto-phoneme for every set of correspondences which cannot be collapsed or explained away as being due to metathesis, dissimilation, assimilation, dialect borrowing, analogy or the myriad of other exceptions to the Neogrammarian hypothesis of exceptionless sound change, one still ends up with a system which has vastly more phonemes than any AN language. An example of this is the system Tsuchida (1976) uses in his reconstruction of Proto-Tsouic phonology. The phoneme inventory he uses is i, ə, a, u, p, b, m, t, T, C, j, n, ñ, k, g, ŋ, l, R, S₁, S₂, S₃, S₄, S₅, S₆, x₁, x₂, X, Q₁, Q₂, y, s, θ₁, H₁, H₂, N, L, w₁, w₂, W₁, W₂, C₁, t₁, D₁, D₂, D₃, D₄, z₁, j₁, ?₁, ?₂, ?₃, ?₄, K₁, K₂, l₁, rd, q, c, R₁, R₂, R₃, R₄, R₅. Tsuchida himself says that this number of proto-phonemes seems excessive and he states that he hopes that some of them, especially the ones separated by subscripts, will be ultimately unifiable. However, within the approach Tsuchida is using, one is forced to posit these proto-phonemes (and probably on examining more languages, a great many more). I would argue that the problems which have arisen in attempting to reconstruct the PAN phonological system have arisen not because of problems in the data but because of fundamental problems in the approach which has been taken to phonological reconstruction.

The Neogrammarian (hereafter N.H.) hypothesis of exceptionless sound change has never been observed as occurring in reality. As long ago as 1905 Gauchat (1905), in his study of sound change in Charmey, a village in the Suisse Romande, showed that sound change does not occur in the manner posited by the N.H., which states that sound change occurs without the knowledge of the speakers of a language as the result of imperceptible minute variations in the acoustic production of speech sounds.

In his study of palatalisation of *l* to *j* in Charmey, Gauchat found that generation I (the oldest) always had *l*^j and generation III always had *j* but generation II did not have some intermediate phonetic stage as the N.H. would predict. They had instead both *l*^j and *j* in fluctuation.

Labov, who has been the leading force in observing sound change in progress in recent times, has recorded essentially the same phenomena in his study of the centralisation of the diphthong *aw* in Martha's Vineyard (Labov 1963). Labov has shown in other studies of sound change in progress that not only does age affect sound change but other social factors such as class (his study of the reintroduction of *r* into the New York dialect: Labov 1972) and sex (his study of the raising of *æh* in New York: Labov 1966) also affect sound change. Studies of change in other English dialects have confirmed the fact that sound change progresses at different rates according to these social factors of age, class and sex (see Trudgill's study of vowels in Norwich English: Trudgill 1971).

However, this does not mean that sound change occurs at random. Both Labov and Gauchat found that sound change was favoured by particular environments. Furthermore, over time the usage of the variants tends to become more and more consistent. Either one variant will win out over the other or they will be used in different environments. (e.g. Gauchat's generation III speakers always use *j* and Labov's generation III on Martha's Vineyard have a consistent environment for the centralisation of *aw*).

Other factors may interfere with this development towards consistency. Labov found in his study of the raising of *æ* in New York that this pronunciation was a stigmatised one and that people who raised the *æ* would sporadically correct it in situations where they were being careful in their speech. It is well known that the prestige of the dialect, class or person who makes an innovation in the

phonological system has great influence on how far it spreads. Furthermore, even though there is a strong tendency for consistency to develop, it does not always do so. In his analysis of the development of Middle English a:, ε: and e: into Modern English (Labov 1974), Labov regards the development of 'steak' as irregular. In the rotacisation of medial s to r in Latin the retention of s in *nasus* is irregular (Anttila 1972). These irregularities are not due to social or other types of interference but simply to the failure of the sound change to become consistent in all words.

Therefore, it appears that the N.H. does not correspond to reality and it fails to predict the development of sound change when it occurs. The N.H. is not only unsound in its relationship to observed reality of sound change, it is also unsound theoretically. The clearest and most incisive attack on the theoretical standing of the N.H. was made by Hugo Schuchardt in his essay "On sound laws: against the Neogrammarians" (Schuchardt 1885). Schuchardt's attack is not a purely negative one: he presents a clear picture of how he views phonological change occurring.

Schuchardt wrote his essay in pre-phonemic times and it must therefore be reconsidered carefully in the light of phonemic theory. The same thing applies to the N.H. as well. Indeed phonemic theory raises immediate problems for the N.H. because phonemes can only change abruptly. A particular sound is either an allophone of one particular phoneme or it is not, but there is no half-way house. It has always been known that certain types of sound change such as metathesis must be abrupt and these types of sound change have always been problematic for the N.H. We will consider the question of the gradualness of sound change later. For the moment we will return to Schuchardt's criticism of the N.H.

The first, most important and probably the most disputed point raised by Schuchardt is that a phonological system is a psychological entity and as such it should not be affected by purely physical factors such as the continual acoustic variations which occur in speech. Labov demonstrates in his article "On the use of the present to explain the past" (Labov 1974) that speakers may consistently differentiate classes of sounds that they regard as one phoneme. As Labov puts it "small differences do exist. To the native speaker they may be worth noticing and may in fact be impossible for him to label; nevertheless, he regularly produces these differences". If phonological systems can ignore differences which are consistently made, why should a random acoustic variation cause it to alter? Schuchardt argues that sound change is therefore as psychologically conditioned as analogy and borrowing are (i.e. that an alteration in a psychological system must be made by psychologically controlled factors). This point of course depends on accepting Sapir's (1949a) view of the phoneme as something real in the mind of the speaker and not Twaddell's distributional definition (Twaddell 1935). What Schuchardt is saying essentially is that while speech varies continually, this in itself does not cause alteration to phonological systems. For this to occur the variation must be pushed by the same things which control the phonological system.

In furtherance of this argument Schuchardt points out that sound change and analogy interact on one another. If they were motivated by entirely different forces, sound change by physical factors and analogy by mental ones, then they would not affect one another because they would be operating on entirely different planes. From this it follows that sound change and analogy have the same motivation by mental factors (as nobody would wish to argue that analogy is motivated by physical factors).

As a part of his general consideration of the relationship between analogy and sound change, Schuchardt attacks the view that analogy causes irregular and arbitrary sound change. His own example of analogy causing regular sound change is the change of all Old Spanish and Portuguese participles ending in *-udo* to *-ido*. Schuchardt here is arguing for grammatical conditioning of sound change, a phenomenon which now has wide acceptance (see Anttila 1972).

Schuchardt's final point on analogy and sound change is that there is no clear dividing line between them. He gives two examples of the development of Italian from Latin to show this. Latin *comite* becomes Italian *conte* (it should be *comte*) and the Latin phrase *non gravis magis leve* becomes Italian *non greve ma lieve* (it should be *non grave ma lieve*). These are both examples of assimilations though the motivations are different. In the first case the motivation is to reduce the number of places of articulation and in the second it is the association of the words in that particular phrase. Yet both are assimilations, changes that make things more like one another for whatever reason. While it is clear that there are differences between the two assimilations, it is equally clear that there are similarities which would be ignored by saying that one is sound change and the other analogy and never the twain shall meet.

If, as Schuchardt argues, sound change and analogy are motivated by the same factors, then sound change will like analogy not proceed to produce consistent results. This, in fact, is what we find in reality. It has long been known that despite what the N.H. said should happen, sound change does not always proceed regularly.

Workers in lexical diffusion theory have provided extensive data confirming this and they have attempted to provide explanations for sound change as they have observed it. The theory of lexical diffusion is discussed in Wang (1969) and further discussed and exemplified by Chen and Wang (1975) and Krishnamurti (1978). Lexical diffusionists take the view that sound change is phonetically abrupt and lexically gradual (Wang 1969:14). In other words sound change spreads gradually across the lexicon and words may have more than one pronunciation. This means that words which were at one time homophonous can acquire different pronunciations over time (Chen and Wang 1975:261). Furthermore, while sound changes tend to apply to all relevant lexical items over time, they can peter out or meet a competing change.

Chen and Wang (1975) and Krishnamurti (1978) provide extensive data which clearly shows that sound change does proceed through the lexicon gradually. Neither Chen and Wang nor Krishnamurti address themselves to the causality of sound change. Chen and Wang argue that the motivations for particular phonemes altering in particular ways are the articulatory and perceptual constraints of speakers (i.e. sound changes occur in particular ways for physical reasons).

When one considers the causality of sound change within a lexical diffusionist view, it is obvious that sound change cannot be caused by acoustic/physical factors. If one accepts that sound change and analogy are motivated by the same factors, then it is comprehensible that sound change can spread gradually. For example, if a language has a number of suffixes marking the dative case and one of these affixes is analogised to replace the others, it will take some time for this process to be completed and it may never reach completion.

In his 1969 article Wang quotes Sturtevant (1917:82) with approval "We have seen that many sound changes are irregular when they first appear and gradually become more and more regular. The reason is that each person who substitutes the new sound for the old in his own pronunciation tends to carry it into new words.

The two processes of spread from word to word and spread from speaker to speaker progress side by side until the new sound has extended to all the words of the language which contained the old sound in the same surroundings". Sturtevant said, in addition "Such a spread of a sound change from word to word closely resembles analogical change; the chief difference is that in analogical change, the association groups are based upon meaning, while in this case the groups are based upon form". (1917:80)

This is exactly what Schuchardt said in his discussion of the relationship between analogy and sound change. Wang points out that from observation, the sound changes that occur are few in kind. Given this fact and the fact that a sound change spreads through the lexicon over time, there is a very good chance that two or more sound changes may operate on the same material to produce irregular results because the rules will operate on different items at different times and therefore produce different results.

Sturtevant essentially agrees with the views of the lexical diffusionists discussed earlier. It is clear therefore that sound change can produce irregular results.

It is not only the lexical spread of sound change which presents problems for the N.H. The geographic spread of sound change has always been exceedingly problematic. The N.H. states that a particular sound change occurs without exceptions but it does so only within a uniform speech community over a certain period of time. This uniform speech community is the 'same dialect' of the Neogrammarians.

However, no such uniform speech community has ever been found, nor is there any reason to imagine, given religion, politics and economics, that any such community has or would ever exist. To get this community Delbrück (1884) had to define it as the normative average of the speech of an individual at one point in time. Quite apart from the fact that this definition ignores things such as individuals controlling social registers, the difference between a regular change in this community and sporadic phonological change of a group of people is for all practical purposes minimal. Furthermore, as Schuchardt points out, even the speech of individuals is subject to variation (i.e. register control, the difference between speech to strangers and intimates). This discussion should not be taken to deny the existence of speech communities, which are as Labov says (1974) orderedly heterogeneous, only the existence of *uniform* speech communities.

This is not the only area where the N.H. fails to correspond to reality. The N.H. claims that sound change proceeds without any conscious knowledge of the speakers. Metathesis, which can sometimes be stated in the same way as any 'ordinary' sound change, has always been problematic for this viewpoint. In Slavic, for example, all sequences of (C)VRC go to (C)RVC (where R represents r or l). Furthermore, even such Neogrammarians as Delbrück (1884) state that certain sound changes spread because they are considered fashionable or prestigious. This would appear to be irreconcilable with the view that sound change is unconscious. Also, as Vennemann (1972) states, people in any speech community are aware of differences between their speech at various points in time, and of the differences based on all the parameters that Labov has studied (Vennemann cites the change of *hw* to *kw* and *xw* in 20 years in Icelandic).

After having raised these points against the N.H., Schuchardt goes on to present his own view of how sound change takes place and regards sound change as occurring in quite the opposite direction from the Neogrammarians. All change in phonological systems occurs initially as a sporadic change. The only difference between 'regular' and 'irregular' sound change is that 'regular' sound

change is 'irregular' sound change which has been generalised to consistency. Schuchardt points out that it is frequently possible to trace the spread of a phonological change from some area of origin as it spreads to more and more speakers. Vennemann (1972) gives an example of this and also of the generalisation of a sound change in the development of $s \rightarrow \int$ which occurs as follows:

- $s \rightarrow \int$ /l Norwegian
- $s \rightarrow \int$ /lrmnw North German
- $s \rightarrow \int$ /lrmnwpt Standard German.

Schuchardt's own example concerns the development of $\bar{e}$ and $\bar{o}$ to *ie* and *uo* in Italian. He shows that it started out when there was a following *i* and *u* respectively. Some dialects still have it in only this environment, but others have spread it in various stages. In some dialects there are only forms such as *vieni*, *buonu*, *buoni* and the like. However, other dialects spread the changes through paradigms so you get forms like *viene* and *buona*. Still other dialects have forms such as *pietra* and *ruota* where the original conditioning factor is completely lost.

Schuchardt points out that his view of sound change is closer to how sound change occurs in reality than is the N.H. which is essentially a philologist's view of how sound change occurs. This is not to say that there are no problems with the alternative view of sound change Schuchardt is proposing. The first is that he gives no example of a truly sporadic change which generalises. All the changes which he discusses as generalising have some initial phonetic or grammatical conditioning. This is in agreement with what Labov and Gauchat found in their studies. Indeed Labov takes great care to point out that while in an overall speech community a sound change may proceed with great irregularity and produce irregular results, if you analyse the speech community along the social parameters Labov uses you will find sound change proceeding with a great deal more regularity, along phonetic and grammatical parameters, within the social parameters.

Therefore it seems that while Schuchardt's view of sound change is closer to reality than the N.H., it does not correspond exactly to reality. Therefore, I think that you would have to modify Schuchardt's statement that all sound change is in origin sporadic to correspond with Labov and Gauchat's observations. The restatement would run something like 'sound change is not in origin consistent but tends to occur more in certain environments than others'.

The second problem with Schuchardt's view is that it cannot explain things such as Grimm's Law where classes of phoneme alter and follow one another in a chain. The N.H. offers no explanation for these phenomena either. Explanations of these phenomena have, I think, been offered and generally accepted and these explanations can be incorporated into the view of sound change espoused by Schuchardt and Labov. The first which explains why classes of phonemes act together as a class, is the concept of features in phonology. There is no reason why any view of sound change should be restricted to dealing with phonemes only. Schuchardt's view is certainly not so restricted.

Martinet's (1952) concept of push and drag chains, whereby when one phoneme or class of phoneme alters, the nearest one will move into the range vacated, is a generally accepted explanation of chain sound changes. It may be added to Schuchardt's viewpoint without any problems.

The third and most major problem with Schuchardt's and with Labov's views and observations of sound change is to explain why sound change generalises.

There is no doubt from Labov and Gauchat's observations that even though sound change may initially be irregular, it does become regular by generalisation.

Andersen (1973) in his article on "Abductive and deductive change" proposes explanations for the genesis and development of sound change. The first type of change Andersen discusses is abductive change. According to Andersen, abductive change occurs when children are presented with phonological data and they analyse it in a different way from their parents. This means that they have different phonological categories from their parents. Andersen goes on to claim that in order to approximate their speech to that of their parents they have rules which produce surface forms corresponding to the accepted norm. These rules are called A rules. Andersen states that these rules may be elaborated and revised throughout the speaker's life.

The second type of change Andersen proposes is deductive change. This type of change involves altering the speaker's surface forms to correspond to the underlying phonological representations. These changes occur abruptly in each word, but across the vocabulary they occur gradually depending on social factors. Old forms will tend to be used with the older generations.

These two types of sound change would certainly go a long way to explaining why sound change generalises. If a reanalysis of the phonological system is involved, then one would expect all instances of the phoneme(s) affected to be altered. However, this is not necessarily the case because as we know that sound change does not always affect all instances of a phoneme (see Labov's (1974) discussion of the Middle English development of *e*: *ɛ*: and *a*:). Furthermore, in his study of his child's acquisition of the English phonological system and its categories, Smith (1973) found that while his son normally produced words which corresponded to the correct English phonological form in terms of the child's own phonological system, there were occasional exceptions when the child had miscategorised something. Therefore, it does not seem that, if children did categorise differently from their parents, they would necessarily do this without exceptions.

Another part of Andersen's hypothesis that would need some testing before being accepted is the A rules. Labov in his study of centralisation does not mention whether the middle generation of speakers in Martha's Vineyard tended not to centralise their vowels when talking to the older speakers. Labov does discuss how New York speakers tended to alter their raised vowels when they were speaking very carefully. They did this sporadically, however, and they did not do it in ordinary speech. This could also be regarded as an example of Andersen's adaptive change (changes due to factors outside the linguistic system). Abductive and deductive change are types of evolutive change (a change explicable in terms of the linguistic system which gave rise to it).

Adaptive change occurs when the native forms of a particular area are replaced as the result of extralinguistic pressure. This would most often involve the replacement of items or pronunciations of less prestigious communities by those of more prestigious communities. This is essentially the same phenomena as Andersen's A rules. These two are random assimilations of less prestigious (children's) speech to that of the more prestigious (parents'). Both adaptive change and A rules are simply examples of socially motivated alterations in a speaker's pronunciation. It is quite conceivable that a particular variant may not be socially unacceptable and so A rules may as such never operate. We have already seen in Wang (1969:15) that words can have variant pronunciations. In the examples he quotes neither variant appears to be socially stigmatised.

Andersen's hypothesis would certainly appear to provide a good explanation for the linguistic behaviour of the youngest generations in Labov's and Gauchat's studies. These were the ones who had the sound changes consistently in particular environments. However, it cannot deal with the linguistic behaviour of the middle generation who were not consistent in their use of changed forms. Until the necessity of A rules for all cases of sound change can be proved, one is entitled to assume there are problems with Andersen's theory in relation to the observed behaviour of the middle generations in Labov's and Gauchat's studies.

Perhaps the most important point about Andersen's view of sound change is that he, like Schuchardt, posits a psychological motivation for sound change. While it may not be able to explain all the observed facets of sound change, it does seem to provide a good explanation for the final stage of sound change as observed by Labov and Gauchat.

As Andersen points out, if you take a view of sound change something like his, then the question of whether sound change is gradual or abrupt loses much of its meaning. Sound change, when it involves abductive change, is neither gradual nor abrupt; deductive and abductive change are phonemically abrupt but across the vocabulary they spread gradually. Basically, according to Andersen, when it is meaningful to talk of gradual vs abrupt, sound change is psychologically abrupt but its linguistic manifestation is gradual.

However, I would argue that unpredictable correspondence sets are a natural consequence of sound change. Wang, Labov and Andersen agree that while sound change starts out being irregular (though associated with conditioning either phonological or grammatical), it will over time produce consistent correspondence sets. However, as Labov demonstrates, this progression towards regularity holds only within particular groups in the speech community. The development of a sound change when the speech community is looked at as a whole can be quite irregular. Once the various component groups of the speech community start to influence one another's speech, as they are continually doing, then a combination of different conditioning factors and different stages of development can cause irregular results. As Schuchardt points out, it is meaningless to call this dialect borrowing. Within a speech community a sound change must originate somewhere and from there it spreads. The social boundaries which are so important in determining how a sound change develops are not absolutely rigid. Therefore, it is natural that the development of a sound change in one section of the community will influence and be influenced by its development in all sections of the community. As the geographical boundaries of speech communities are not rigid either, it makes very little sense to attribute inconsistencies in the result of one sound change to 'borrowing' and then to describe another sound change with consistent results as having spread throughout the same area. The term borrowing should be reserved for those cases where you have languages which are not mutually intelligible or which are geographically separated (i.e. in those cases where you cannot have a common speech community). Borrowing within a speech community is something that is entirely to be expected.

This is essentially what Bloomfield (1933:364) says:

We must suppose that, no matter how minute or accurate our observation, we should always find deviant forms, because from the very outset of a sound change, and during its entire course, and after it is over, the forms of the language are subject to the incessant working of other factors of change, such as, especially, borrowing and analogic combination of new complex forms.

From this discussion of sound change it can be seen that while sound change has a tendency over time to produce consistent results, a number of factors may in the natural course of development of any sound change operate to prevent the sound change producing consistent results.

The view of phonological change, which we have presented, naturally affects the method we use to reconstruct an inventory of proto-phonemes. Because we accept that change can and does produce inconsistent results, we cannot accept all correspondence sets as valid. There is no guarantee that all correspondence sets which are different and which cannot be shown to be conditioned variants do in fact represent different proto-phonemes. Therefore, we will need other factors to determine the acceptability of correspondence sets as evidence for particular proto-phonemes.

Before accepting a correspondence set as evidence for a proto-phoneme we will require that it should occur in a group of cognates which occur throughout Austronesia. If a correspondence set occurs in a set of words which occurs only in one area and not others, then it cannot be accepted as PAN because it could easily have been innovated in that area. We will also require that it should occur in more than one or two words because a correspondence set involving only one or two words could arise by sporadic phonological change.

Before accepting a correspondence set, which occurs only in a small number of non-basic items of vocabulary, as representing a separate proto-phoneme, we will require that the correspondence set have reflexes in a number of geographically separated languages which serve to separate it from other correspondence sets. Otherwise we could again be dealing with an areal innovation.

Taking the PAN phonological system as set up by Dahl (1976) as a starting point, and taking Dempwolff (1934-38) into consideration also, we will consider what changes should be made in the light of the criteria proposed above. The system Dahl sets out (1976:101) is:

Vowels: i u
 ə
 a

*i and *u may be syllabic or non-syllabic.

Consonants:	Stops and affricates		Nasals	Continuants	
	Voiceless	Voiced	Voiced	Voiceless	Voiced
Bilabials	p	b d ₁	m		
	t ₁		n(1)	ʈ	ɭ(1)
Apicals		d ₂		s	r
	t ₂				
	k'	d ₃ d'	n'		
Palatals		ʒ g'			
	t'				
Velars	k	g	ŋ		ɣ
Uvular	q				
Laryngeal				H?	

We agree with what Dahl says (1976:55-69) concerning Dempwolff's *t̥ and *d̥. Dahl says these two phonemes were innovated under Indic influence, in the same way the voiced aspirates were innovated in Madurese. This led to their cognate sets being confused and contradictory. We also agree with the suggestion that the languages where t̥ and d̥ are found originally had a dental t and an alveolar/retroflex d, as is found in some Western Austronesian languages still (Henderson 1965). This situation results from the merger of an original *d and *d̥ (quite separate from Dempwolff's *d̥), and made the languages especially receptive to the introduction of phonemic retroflexes from Indic.

We will now go on to consider Wolff's (1974) proposal to remove *r from the PAN phoneme inventory. He argues that all correspondence sets which are cited for *r can be dealt with as reflexes of either *d or *ɣ. In the Oceanic languages *d and *r are always reflexed as r. Indeed, according to Wolff it is only TB and M1 which require us to establish *r as a separate phoneme. Furthermore, Wolff states that he can find only two cognates for r in Taiwan. However, Tsuchida (1976:177, 227) provides several cognates for *r on Taiwan, five of which I find convincing, and which, I think, compel us to reconstruct *r. The cognates are:

1. *garuts Pw garuts 'comb', M1 garut 'scrape', Sa keru 'to scratch'.
2. *kurap Sar kurapə 'scabies', Pw kurap 'a scaly skin disease', M1 kurap 'ringworm', Tausug kulap 'scabies'.
3. *buar Kn ma:-vu:-vuara 'to dissemble after a meeting', Sar m-i:-vuara 'to collapse, tumble down on piles of firewood, stone', Tg buwal 'fall, fell, demolish, drop, fall on ground'.
4. *kuruŋ Pw kuruŋ 'wooden frame for carrying firewood', Ts t-h-i-kruŋu 'surrounding walls of a pig pen', Ts re-kruŋv-a no hopo 'mosquito net', Tg kulón, Ilokano kurún, Bikol kurún 'fortress', Tausug ku:ŋan 'cage', To kolo 'fortress, temporary fence of native cloth around a grave', Sm ŋolo 'fort'.
5. *gərgər Pw m-igərgər, Kan m-a:-kərkər, Sar k-um-a-kə:rə kə:rə, Ts s-m-o-ʔərʔərə, Ilokano pi-gərgər, Kampampangan galgal, Proto-Minahasan *gərgər M1 gəgar 'to tremble'.

In Paiwan, Saaroa, Kanakanabu none of the other voiced apicals *d, *l or *d̥ ever has an r reflex. In Tsou *d̥ gives r but it is the only voiced apical to do so. In all the Tsouic languages *ɣ gives r; in Paiwan *ɣ gives ø normally but r occasionally. In Tg, Bikol and Tausug *ɣ gives g. In Kampampangan *ɣ gives y.

Taking the cognates one at a time, we will show that *r is the proto-phoneme to be reconstructed rather than *ɣ or a voiced apical. For the first cognate we must note that the Paiwan evidence shows that this is derived from gutsguts 'to scratch' and the infix ar (having sound or quality of; involving the use of). Now this -ar- in Paiwan could come from *-aɣ- or *-ar-. However in Tg these two are reflected as -ag- and -al- being basically verbalising and adjectivalising infixes respectively (Zorc, personal communication). The Paiwan -ar- is obviously cognate to -al- not -ag-. Therefore we have to reconstruct *-ar-. An infix with *r is extremely strong evidence for the phoneme's existence.

The second cognate is not quite so convincing: the Pw, M1 and Sar forms could come from *kuɣap but the Tausug form cannot. The Tausug form could be a Malay borrowing as it is the only Philippine cognate. Therefore *kuɣap is not too certain. The third cognate is much stronger: the Kn and Sar forms cannot

reflect a voiced apical and Tg cannot reflect *ɣ and there is no Malay form so Tg cannot be a borrowing. The fourth cognate is equally strong. The Paiwan form excludes a voiced apical. The Tg, Bikol and Tausug forms exclude *ɣ. The Philippine forms do not appear to be borrowings. The Tg, Bikol and Ilokano forms all have final accent whereas if the borrowing was from Malay one would expect at least some forms with initial accent (Zorc, personal communication). The Tausug form shows the loss of l (*r→l in Tausug) between similar vowels which is at least 1,000 years old (Zorc, personal communication). The fifth form provides perhaps the strongest Philippine evidence. Pw, Kn, Sar and Ilokano all exclude a voiced apical (r cannot be a reflex of a PPh voiced apical in this position (Zorc, personal communication). Kampampangan excludes *ɣ. The Philippine forms cannot be borrowed from Malay as the Malay reduction of *gəgəg to *gəgə is very ancient.

Therefore it appears that we must reconstruct *r. However, we should note that *r does not occur initially. This is significant in light of the fact that a number of Austronesian languages have r as the intervocalic allophone of d. The difficulty of finding cognates for *r and its defective distribution both suggest that it was a very rare phoneme of recent origin in PAN.

Though most proposals for alteration of the system originally proposed by Dempwolff have involved reducing the number of phonemes, there has been one set of proposals to increase the number of phonemes which is generally accepted. These proposals arise because of Taiwanese data which was not available to Dempwolff. The new phonemes are *ts, *d₂, *ɖ (different from Dempwolff's *ɖ) and *ɣ. Outside Taiwan *ts merges with *t, *d₂ and *ɖ with *d, and *ɣ with *n and *l. The correspondence sets for these phonemes involve such basic and widespread words as 'eye', 'excrement', 'ear', 'die', 'water', 'black', 'two', 'hear', 'child', 'swim' (see Dahl 1976:55-59).

The alternative to accepting these new proto-phonemes is to regard them as unconditioned areal splits (*ts being *t, *d₂ and *ɖ from *d and *ɣ from *l and *n). Each of these possible splits will be considered. The idea of initial ɣ being the result of an unconditioned split from *l and non-initial *ɣ being an unconditioned split from *n completely lacks credibility; if indeed it is possible that such a sound change could occur. Therefore *ɣ must be accepted.

The idea of a double split of *ɖ and *d₂ from *d also lacks credibility. We will consider *ɖ first. In Amis, Tsou, Thao, Favorlang, Paiwan and Siraya *ɖ is reflexed separately from *d (see Dahl 1976:64). In Amis and Thao *ɖ is reflexed as t and *d as r and ø respectively. In Tsou, Favorlang and Siraya it is r whereas *d is ts, t and d respectively. Too much reliance should not be placed on Favorlang and Siraya owing to the paucity of cognates and uncertainties as information is available only from 17th century Dutch manuscripts. Nevertheless, it does appear from Marsh's (1977:76-79) study of Favorlang that *d and *ɖ have different reflexes. The reflexes of *d and *ɖ are sufficiently different and consistent for us to say that there appears to be enough evidence for *ɖ. Also as Dahl points out the existence of a PAN *ɖ would help to explain the fact that t is dental and d is alveolar in certain western AN languages (Henderson 1965). We would have a merger of *d and *ɖ to ɖ while *t would remain dental.

The evidence for *d₂ is very much weaker than that for *ɖ. The main evidence comes from Paiwan (see Dahl 1976:59), with support to some extent from Puyuma, Favorlang, Siraya and Kavalan. Of these the Puyuma evidence appears to be unreliable, and even if it is reliable, the reflexes in Puyuma of the voiced apicals are too varied to form a clear basis for separating *d₁ and *d₂.

Favorlang appears to have merged $*d_2$ and $*d$ (see Marsh 1977:59-61, 76-79). However, as we pointed out earlier the material available on Favorlang is not reliable enough to be a base for positing a distinction in PAN of $*d_1$ vs $*d_2$. With Kuvalan and Siraya there are simply not enough cognates to support the distinction for PAN. Siraya has the following $*d_2$ cognates salom $*d_2a\text{um}$ 'water', ricos $*likud_2$ 'back' and houkas $*t'ukud_2$ 'stick' [found only in Taiwan]. Siraya has d for $*d_1$. Furthermore Siraya has the same problems as Favorlang. The only sources we have on the language are 17th century Dutch sources which are, as we have said, not a reliable enough base for a PAN proto-phoneme distinction. Kuvalan has only one $*d_2$ cognate ranum $*d_2a\text{um}$ 'water'.

This is not sufficient evidence to discount the possibility of an areal innovation, given that the reflexes of $*d_1$ and $*d_2$ are phonetically similar. There are variations in the reflexes of $*d_1$ in several Taiwanese languages (Dahl 1976:64). Therefore unless more evidence becomes available from a more detailed study of the Taiwanese languages, we must regard the case for $*d_2$ as unproven.

On the other hand, $*ts$ is reflexed separately from $*t$ by all the Taiwanese languages except Amis, Bunun, Siraya and Kavalan. Furthermore, there are extensive cognate sets to attest the difference (Dahl 1976:61-64). Furthermore, $*ts$ is fairly consistently reflexed. If $*ts$ was an unconditioned areal split from $*t$, we would expect there to be variation among the languages (e.g. $*matsa$ being reflected as $*mata$ in some languages which were being used as criteria for separating $*ts$ from $*t$). Therefore, the evidence appears good enough to accept $*ts$.

We have now dealt with the apicals and we will now go on to consider the palatals. This is the most problematic area in reconstructing the PAN phonological system; the area where the correspondence sets are least consistent. The first set of palatal phonemes we will consider are Dempwolff's $*d'$ and Dyen's $*Z$ (Dyen 1951). These two phonemes are obviously very similar in nature as they have the same reflexes with different distributions.

The two languages which require us to reconstruct a $*d'$ separate from $*Z$ are Javanese and Toba Batak. The Javanese and Toba Batak reflex of $*d'$ is d' and of $*Z$ is d , but M1 shows d' for both. However, if we examine the cognates we find that in Javanese and Toba Batak d is the reflex you get in basic words such as 'road' $*Zalan$, 'rain' $*quZa$ (two of the most widespread cognates throughout Austronesia), 'tongue' $*Zilaq$, 'compass-needle' $*Zayum$ (the Austro-nesians were and are a seafaring people). In Javanese there are at least two cognate sets which have inherited Javanese words with d and Malay loans with d' . There are d' arum 'mediator' (must be from Malay because $*y$ gives ϕ in Jv and r in Malay) vs dom 'needle' (Native Javanese); and d' uruh 'syrup' (must be a Malay loan for the same reasons as before) vs du-doh 'sap'. In his discussion of $*Z$ Dahl (1976) discards two Toba Batak words pod'am 'close the eyes' and in'd'am 'borrow' (this has a cognate in Paiwan) as Malay loans. We know from Dyen's (1956) word on Ngadju Dayak that Malay can have an immense influence on the vocabulary of languages. In Ngadju Dayak the Malay loans form a greater percentage of the vocabulary than the inherited forms. In this case Dyen states that to determine the native reflex one must accept the reflexes given by the basic words. We know from inscriptions that Malay has been known in Java for 1,000 years and its influence has steadily increased through time. Furthermore, Madurese appears to have d' or $d'h$ as its reflex of $*Z$. It appears therefore that there is good reason to say that Javanese and Toba Batak d' have come from Malay or Madurese loans. This would mean that there is no necessity for reconstructing $*d'$ as a phoneme separate from $*Z$. There is another very good reason

not to reconstruct *d' as a phoneme separate from *Z. There are only two cognates for *d' on Formosa. One is a plant name (admittedly an important one, the betel-nut) and the other a word for 'sharp'.

Jv d'ambe Ml d'ambi 'areca palm', Md d'ambe 'a vegetable', Pw qa-zavai 'Machilus Kusanoi'.

Amis si-tarum, Ss s-əm-aəm, TB tad'om, Jv tad'əm, Ml tad'am, NgD tad'im 'sharp'.

The cognate for 'sharp' raises some problems for our argument that the Javanese and TB forms with d' are Malay loans. The Jv and TB forms have the native ə and o reflexes of *ə as opposed to the Ml a reflex. There are two possible answers to this problem. The first is that the form was borrowed from Malay when *ə was still ə in all three languages. The other is that the presence of d' in Jv and TB results from the spread of the phoneme once it was established in Jv and TB by Ml loans. This would be a type of hypercorrection. Exactly the same phenomenon occurred in Madurese with the voiced aspirates which were established by Indic loans (e.g. *damay Md dhamar). If Md can replace d with dh, it is perfectly possible that Jv and TB have replaced d by d'. I prefer this explanation to the first one proposed, because it correlates with what we know happened in Madurese and with what has happened elsewhere in the world, whereas there can be no proof that what is proposed by the first explanation ever happened.

Dahl and apparently Dempwolff (Dahl 1976) suspect that *d' was not an original PAN phoneme. Unlike most other PAN consonant phonemes it occurs only initially and medially. As we have examples of its spread now, we could reasonably suspect that it has spread since PAN times and that at one stage it was not a separate phoneme, though it certainly must be reconstructed as such for PAN. A possible explanation for the spread *d' is suggested by the fact that Malay has been in contact with the Austroasiatic languages, which have a full series of palatals in all positions of the word, for a long time. Since Emeneau's article on India as a linguistic area (1956) it has been well known that certain classes of phonemes can spread throughout an area without actually being borrowed. As Malay already had d', coming into contact with the Austroasiatic languages would provide an impetus for the spread of d'. Once it had spread in Malay, it could spread in other AN languages, because of the use of Malay as a trade language throughout Indonesia.

We will now go on to consider the status of *n' as a PAN phoneme. We know from Blust's (1978) work that this phoneme occurred in the ancestor language of the Oceanic languages. It occurs in a number of Western Indonesian languages such as Javanese. It occurs in Kapampangan and Chamorro and so must be reconstructed for PPh (Blust and Zorc, personal communication). However, in Formosa I have been able to find only three cognates with *n' from Tsuchida (1976:229). They are as follows:

1. Pw sə-qaʔud' 'to be carried away by current', Kn m-a-ʔatsunu, Sar m-u-aʔusa, To ŋ-ohtsu, Ruk mu-aʔuʔu, mu-luʔu, mu-ʔuʔu 'to be adrift', Bun maŋ-hanuʔ, Ss lö-ʔälur, Pl mu-lahuʔ 'flow', Tg ʔanod, Hl Bs anod, Ml haʔut, Jv aʔut, To mu-ʔanu 'float' (Am ma-qalul, Se qəlul, Ruk ʔarur 'point to a form' *qalul)
2. Pw qa-ʔits 'skin, pod shell', Kn ʔaŋi:tsi, Sar ʔaʔitsi, Ts hitsi 'leather', Kel anit 'skin, bark', Tg anit 'scalp', WBM anit 'leather', Bk anit 'animal hide', Tebeduh Piching aʔit 'skin'.
3. Kn ʔa:la 'flame', Thao ʃu-na:raʔ 'to burn', Tg nala 'flame', Sa nala 'to explode, as steam'.

As can be seen Kn is the only language with reflexes in all three forms. It has two separate reflexes; η and n , which are not conditioned in any obvious way. If we examine Tsuchida's material (1976:229-230) we see that Kn η corresponds to Sar n and η (which in turn correspond to $*n$ and $*\eta$). We know that in Kn $*\eta$ and $*n$ have fallen together as n . Therefore it seems that the best explanation for the η in Kn, where we would expect n , is that n has sporadically changed to η in Kn. Therefore, we can regard the Kn reflex in all three cases as n . When we examine the other Formosan forms we see they point to forms $*qa\eta ud$, $*qa\eta its$ and $*nala$. In Thao the reflex of $*n$ is n and $*\eta$ is η or l ($*ini$ 'this', Th $ináy$). Therefore Thao $\int u-na:ra$ indicates a form $*nala$ rather than $*\eta ala$. Paiwan reflects $*\eta$ as η and $*n$ as n ($*panaq$ 'arrow', Pw $panaq$, $*da\eta um$, Pw $za\eta um$ 'water'). Therefore Paiwan indicates $*qa\eta ud$ and $*qa\eta its$. Unfortunately we do not have a Paiwan reflex for $*nala$, nor do we have any from any of the other Formosan languages which distinguish $*n$ and $*\eta$ — in fact all the languages except Kanakanabu, Bunun, Puyuma and Kavalan (Dahl 1976). So far we have been working on the basis of $*n'$ and $*\eta$ merging to η in Formosa. Now we are faced with the possibility that n' is reflected as $*n$ and $*\eta$. While in our view of phonological change an unconditioned split of a phoneme is not impossible, other solutions to the data should be considered. It could be argued from the forms $*nala$, $*qa\eta ud$ and $*qa\eta its$ that $*n'$ has split to $*n$ initially and $*\eta$ medially. Against this one may raise evidence from Paiwan that $*qa\eta its$ consists of a prefix qa and a base $*\eta its$ (Ferrell 1982). Therefore if $*n'$ merges with $*n$ initially we would expect the Formosan languages to reflect $*qanits$ (unless one maintains that the split merger took place after $*n'its$ ceased to be used on its own — in Paiwan it appears that its does not occur on its own).

Furthermore $*qa\eta ud$ - $*qan'ud$ is a somewhat unstable base to build a proto-phoneme on. As pointed out in the list of cognates Amis $na-qalul$, Se $qelul$, Ruk $arur$ all unambiguously point to a form $*qelul$ 'flow'. In Amis $*qan'ud$ should reflect as $qa\eta ur$, in Sediq as $qalud$ and Ruk as $\eta alud$. This means that $*qa\eta ud$ - $*qan'ud$, the most widespread cognate is subject to doubling; a process which Ferrell (1982) describes as quite extensive in Paiwan. Therefore, this means that the Formosan forms and the non-Formosan forms could descend from different though phonologically similar doublets.

It appears that the reflexes of $*n'$ are very problematic and full of doubts for the few cognates which can be found on Formosa. We will therefore consider alternative solutions. The most obvious of these is to eliminate $*n'$ from the PAN phonemic inventory. If we do this then we are saying that $*n'$ arose at some period after the break-up of PAN. In section 3. a mechanism which would have caused this is discussed. Once n' was in the phonemic inventory it could easily be extended to forms with n (i.e. n becomes n' sporadically). Outside Formosa $*n$ and medial and final $*\eta$ are always n and so with the cognates we have $*qa\eta ud$, $*qa\eta its$ (there is no evidence that outside Formosa $qa\eta its$ was an analysable form) $*nala$ would be reflexed as $qanud$, $qanit$ and $nala$ which could easily become $qan'ud$, $qan'it$ and $n'ala$. There is other evidence for alternations between n and n' . In his discussion of the POC palatals Blust points out that the word $n'a[]$ or $na[]$ ($[]$ indicates that another syllable is usually added) which indicates 'mother' and various female relatives must be reconstructed with n' and n because of contradictory evidence. Minangkabau and Chamorro which usually reflex $*n'$ reflex this item with n , whereas Wogeo and some Admiralty languages reflex it with n' . Another case where Wogeo has a n' and a western language which usually reflexes $*n'$ as n' — in this case Malay — has n is tuna Ml, tun'a Wogeo 'eel'. Much the same as this example is the correspondence between Barepe $n'opu$ and Chamorro $nufuq$ (from $*nop(u)$ 'stonefish'). To solve the problems

with correspondence sets involving the POC root *natu 'child, offspring, ovary, egg' which has reflexes indicating a form n'atu, Blust proposes two roots *natu 'child' and *n'atu 'egg, ovary, baby bird'. With the view of *n' proposed here, the pressure to do this is lessened and while there may be other reasons to propose the two roots (and Blust does provide other good reasons), this is possibly another case of n' and n interacting. Finally there are two cases of n becoming n' when followed by i. The word for 'coconut' *niuy shows n' reflexes in many languages and the 3sg genitive *nia is reflexed as n'a in many languages. Therefore we have a situation where n' occurs when one would expect n, and where it appears that n' may correspond to two different phonemes in Formosa, which are reflexes as n outside Formosa in the cognates we have. These facts clearly argue that *n' arose after the break-up of PAN and in section 3. we will provide a mechanism for its becoming a phoneme. Therefore we do not reconstruct *n' as a PAN phoneme. This non-reconstruction accords with Dahl's and apparently Dempwolff's feeling that it did not occur in PAN.

Another of the phonemes which Dahl and Dempwolff suspected was not of PAN status is *k'. The only cognate Dahl cites for *k' on Taiwan is *k'aiŋ Pw tsaŋ 'to tether, to tie', Tg sahiŋ 'tar', Jv t'eŋ 'syrup', NgD saiŋ 'together'. However, as Tsuchida (1976:186 note 3) points out, Tg sahiŋ is from PAN t'aləŋ 'pine-tree, resin', so we can remove Tg. This leaves three forms with an extremely dubious semantic connection. It appears that we cannot accept these three forms as a basis for *k' cognate in Taiwan. There is however a *k' cognate in Taiwan (Zorc, personal communication). This is *pək'əq 'shattered, broken' Pw pətəq, Northern Tagbanwa pəsək, Jv pəcah, Ml pəcah, Fj voza, Sa pota. However, this is the only Taiwanese cognate known.

We should also consider some other facts about *k'. Outside a group of Western Indonesian languages (Javanese, Malay, Acehnese) and Malagasy, *k' always has the same reflex as *t'. In Malay, Javanese, Acehnese and the other Western Indonesian languages it has the reflex t' and in Malagasy it has ts (presumably from an earlier *t'). This is highly suggestive that what we are dealing with is not a separate phoneme but a sporadic retention of *t' in Western Indonesia. (It is generally accepted that the Malagasys came from Southern Borneo). Furthermore, we can provide a good reason why *t' should be sporadically retained as t'. Once *t' started to develop into s, a development which occurs throughout Indonesia and the Philippines, an unbalanced system of *d', *n' but no *t' would be left. To remedy this there are two solutions, either lose *d', as occurs in the Philippines, or sporadically retain *t' as t' in the non-basic vocabulary (an examination of the *k' cognates listed in Dempwolff 1938 shows that none of them occur in what could be called the basic vocabulary). As we have already mentioned, Malay, which we know has exerted a strong influence throughout this region, is in contact with the Austroasiatic languages which have a full palatal series in all positions. Therefore we do not accept *k' as a PAN proto-phoneme. To do so would be to set up a PAN proto-phoneme on the basis of a small number of non-basic vocabulary items which are largely confined to one area.

Therefore we have the following palatal phonemes *t', *d' and *g'. We should make some attempt to determine the phonetic nature of these phonemes. There is little doubt that *d' was d'; it has reflexes d, d' and s. *t' has reflexes t (Pw, Ruk, Th, Pl, Motu and Sa), s (most languages), ts (Amis), and dz (Pz). In the AN languages where nasal substitution occurs n' is the nasal which substitutes for *t' (if the language has n', otherwise n substitutes). (See section 3.). This argues that *t' was indeed a voiceless palatal stop.

The realisation of *g' is much harder to determine. It has reflexes d, ɣ, s, g, l, r, dz, n. As Dahl points out this tends to suggest a voiced palatal

affricate. However, this is unlikely because this would sound very like *d' which seems fairly clearly to have been d'. *g' could indeed have been a palatalised g or it could also have been a palatal fricative. A palatal fricative would be parallel to the apical s and velar γ. Phonological patterns are not however a particularly good reason to reconstruct a phoneme in a particular way. The reflexes are really too varied for the phonetics of *g to be reconstructed so we will continue to use the symbol *g'. This implies nothing about its actual phonetic realisation.

The final alteration which I would like to propose to the PAN phoneme inventory set up by Dempwolff is to consider the place of *g as a separate phoneme. The only cognates involving g that I am aware of in Formosa are gəmgəm ('fist, hit with fist' Dempwolff) git'git' ('cut into pieces' Pw gigit 'bite' Dempwolff) garuts ('comb' Pw, garut 'scratch' elsewhere) gatsəl ('itch' Pw, gatəl 'itch' elsewhere). However, garuts Paiwan is obviously derived from gutsguts 'scratch' and the infix ar 'having sound or quality of, involving use of'. Therefore, the word Dempwolff reconstructed as garut must be a frozen derived form from an earlier verb gut. Otherwise the Paiwan form could not be cognate. Therefore, we have three reduplicated forms *gəmgəm, *git'git', and a normal disyllabic root *gatsəl.

In the Philippines these forms all have doublets with k (Zorc, personal communication). This immediately raises questions as to their validity. Furthermore these roots all describe actions which are frequently described with onomatopoeic words, e.g. their meanings are 'hit with a fist, fist', 'cut into pieces, bite', 'scratch', and 'itch'.

There are a number of points about *g raised by Dahl (1976:86) which need to be considered. Firstly Dahl says that there are no instances of *g and *γ in the same wordbase. However, this is not correct, there is a root *tagəγaŋ 'ribs' reflected in Taiwan, the Philippines and Borneo (Blust, personal communication). This form has a k doublet (Zorc, personal communication) though. Dahl goes on to point out that γ occurs in only one reduplicated stem piγpiγ/miγmiγ 'spout' which has no Formosan cognates. On the other hand, Dahl points out that there are many instances of g in the same wordbase and not only in wordbases of the structure CVC².

Therefore it seems that *g has some sort of interaction with both *γ and *k. The question naturally arises as to whether there was a *g in PAN. The doublets with *k immediately suggest that *g is an innovation. However, there are no doublets in Paiwan, so one could equally validly maintain that the k doublets arose later in the Philippine languages. The occurrence of *g in words which are similar semantically, all having an onomatopoeic value and three out of four of which are CVC², and the non-occurrence of *γ in CVC² words makes one suspect that maybe *g and *γ should be combined. Against this we have the fact that *g and *γ are separately reflexes everywhere except in a few Philippine languages. Also we have *gatsəl, an ordinary disyllabic root.

From this evidence it appears that we must accept *g as a phoneme of limited occurrence with a phonaesthetic value. The status of *g in PAN would be analogous to the status of ŋ in present day English. *g was probably a fairly recent innovation as a phoneme in PAN, probably from *γ, it being the form of *γ used in CVC² words involving the idea of 'cut, rip, scratch' and then spreading from that start. It is also possible that it was the form of *k used in such words. Whatever its origins, it appears that *g was a PAN phoneme.

The last proposal concerning the PAN consonant inventory which we will consider is the one made by Dahl concerning *w and *j. Dahl demonstrates convincingly that *w and *u, and *i and *j were in complementary distribution in PAN. He does this by eliminating all sequences of *ij, *ji, *uw and *wu from Dempwolff's reconstruction. He shows that hiatus was the PAN thing found between combinations Vi, iV, Vu, uV. The languages which have j and w in these environments have inserted these glides. (There is usually no correspondence between words in the languages which insert the glides — a word with w in Jv does not usually have a cognate with w in Tg).

However, as Blust (1976) points out, the same complementary distribution applies in many present-day Austronesian languages but there are reasons to regard the non-syllabic i and u (j and w) as glides. Frequently failing to do so causes great difficulty in making statements about phonotactics and phonological processes in many languages. Furthermore, for example, Amblau has treated final j as a consonant in that it is lost like all other final consonants whereas final vowels are not lost. As Blust points out, adopting Dahl's proposal would obscure the dominant disyllabic character of the PAN lexicon. Therefore, I think until further evidence is uncovered the best solution is to regard *i and *j, *u and *w as being in complementary distribution but that there are phonotactic reasons for regarding *j and *w as glides.

Therefore the PAN consonant inventory we are proposing is:

p	t	ts	t'	k	q
b	d	d̥	d'	g'	g
m	n				ŋ
	r				
	l, ɭ				
	s			y	h
w			j		

2. PAN GRAMMAR

2.1. PAN syntax and verbal morphology

There are three main syntactic types in the Austronesian languages. A focus system is found in the Philippines and Formosa; this is frequently called 'Philippine type' syntax. The two other types of syntax are 'Oceanic type' and 'Indonesian type' (Javanese, Malay). Philippine and Oceanic syntax are very different whereas Indonesian shares the features of both to a greater or lesser extent. The problem in reconstructing the clause structure of PAN has been which of these three syntactic types to reconstruct.

Foley (1976) argued that one should reconstruct PAN clause structure as being that of the Oceanic languages. Wolff (1973) appears to argue for the reconstruction of a pure Philippine style focus system. Pawley and Reid (1976) argue that PAN looked something like the Indonesian languages such as Toba Batak. Of these three approaches I consider Pawley and Reid's to be the one which explains the data best and which involves the most plausible changes.

Before discussing Pawley and Reid's proposal we will list the major properties of the Philippine and Oceanic clause structures within the framework of Role and Reference Grammar. Pawley and Reid present a description of POC clause structure which is generally accepted (Foley 1976). Therefore we will use this reconstruction, but rephrase it in terms of Role and Reference Grammar. The following are major features of POC clause structure.

- a) In transitive clauses Actor Verb Object is the unmarked word order.
- b) The verb consists of a stem and a clitic Actor concord pronoun and a tense/aspect marker which precede the verb stem. If the verb is transitive then, following the verb, there is a transitive suffix and an embedded pronoun which agrees with the Object.
- c) Actor and Object NPs have case marking whereas obliques occur as prepositional phrases.
- d) Actors are usually restricted to being agents, forces or experiencers.
- e) There are two transitive suffixes *i and *aki(ni). When *i occurs, this marks that the Object is a patient or product with an agentive verb, a stimulus or target with a psychological verb, and a location or goal with a verb of motion or posture. The transitive suffix *aki(ni) marks that the Object is an instrument with an agentive verb, a comitative with a motion or posture verb, and a causal or comitative with psychological verbs. It also marks that the Object is a benefactive.
- f) *i and *aki(ni) also occur as prepositions introducing obliques.
- g) Verb roots are divided into stative and active classes. Active verbs are divided into two classes; Action verbs (A verbs) and Process verbs (P verbs) (Arms 1974 - his analysis is of Fijian but it applies to POC). These verbs, when they occur as bare stems form intransitive clauses. To form transitive clauses, one can add the transitive suffixes *i and *aki(ni) to the stem. One can also form transitive clauses by adding the causative prefix *paka and *i or *aki. When an A verb is transitive the agent role, which occurs when it is intransitive, is the Actor. When a P verb is transitive the patient role, which occurs when it is intransitive, is the Object.

So far we have been using the term Object which is not a term in RRG without justification for so doing. The reason we are using the term 'Object' is that we need some term to cover the non-Actor nuclear role which occurs in transitive clauses in POC. Our use of the term implies nothing more than this. It is obvious that there are two types of Object, those marked by *i and those marked by *aki(ni). It appears that in POC all verb stems were monovalent and to make them bivalent, one had to add the transitive suffixes.

There are no such restrictions on the valency of verb stems in the focus languages. This is a correlate of the fact that clause structure is quite different in the focus languages from POC. Just how different it is can be seen from the following examples which show the four ways one can say in Paiwan 'the man hunts the pigs in the mountains with a spear'.

- | | |
|----------|--|
| Actor | 1. q/m/aʔup a tsautsau tau vavuy i gadu tua vuʔuq |
| Focus | hunt man pigs mountains spear |
| | 'the man hunts the pigs in the mountains with a spear' |
| Patient | 2. qaʔup- n nua tsautsau a vavuy i gadu tua vuʔuq |
| Focus | 'the pigs are hunted by the man in the mountains with a spear' |
| Locative | 3. qa up-an nua tsautsau tua vavuy a gadu tua vuʔuq |
| Focus | 'the mountains are where the man hunted the pigs with a spear' |

Instrumental 4. si-qaɬup nua tsautsau tua vavuy i gadu a vuɭuq
 Focus 'the spear is what the man hunted the pigs in the mountains
 with'

These four focus types are found in almost all focus languages and can be reconstructed for the Proto-Focus Language (PFL). The English glosses of these Paiwan clauses are not accurate reflections of the Paiwan. The English glosses imply that the Actor Focus (AF) construction is the basic construction. This is not the case with the Paiwan clauses. As can be seen from the clauses, the verb is equally marked in each case. This marking indicates that either the Actor is in focus, or the patient (PF), or the locative (LF), or the instrumental (IF). The NP in focus is indicated by a. None of these constructions can be said to be basic. They are all equally possible candidates for the basic structure. One can equally possibly derive the structure of the AF, PF and LF clauses taking the IF clause as basic, as one could derive the PF, LF and IF clauses taking the AF clause as basic, and so on. The NP in focus is the pragmatic peak (PrP) of the clause. The pragmatic peak is the centre of attention for the clause. In an IF clause, the hearer's attention is directed by the speaker to the instrument as being the most central participant role which is to be the pivot for the hearer's attention. In the focus languages any NP has an equal chance of being PrP with any other NP because of the four focus types. As a result of this system of clause structure, the properties associated with subjects (Keenan 1976) are split between the PrP and the Actor (Schachter 1977). The PrP controls referential properties such as relativisation, coreferential deletion and quantifier float, while the Actor controls role related properties such as reflexivisation (Schachter 1977).

These four clause structures, which occur in the focus languages, also occur in the perfective marked by an affix which is reconstructible for PFL. The perfective counterpart of the Paiwan IF clause given is:

s-in-i-qaɬup nua tsautsau tua vavuy i gadu tua vuɭuq
 'The man hunted the pigs in the mountains with a spear'.

However, though the focus system is very different from the POC clause structure, there are some common elements. The focus languages, like POC, have a class of stative verbs. In the focus languages and POC these verbs are marked by the prefix *ma-*. Some examples from Tagalog are *mabingi* 'become deaf', *mabasag* 'get broken', *mamatay* 'die'. Also found throughout Austronesia are the causative prefix **pa-~paka-* and the reflexive/reciprocal **may-*.

The Paiwan and Tagalog examples given here are representative of what can be reconstructed for the PFL and they show its major characteristics. These features still characterise almost all the focus languages and they are as follows.

- a) The basic word order is V NP*.
- b) One of the NPs is marked as PrP. Its case role is indicated by an affix on the verb. The other NPs normally have prepositioned role markers.
- c) Any nuclear role may become PrP.
- d) There are four focus types, Actor Focus, Patient Focus, Locative Focus, and Instrumental Focus.
- e) These four focuses may occur in the perfective when they are marked by **-in-*.
- f) There is a class of stative verbs marked by **ma-*.

Finally it should be mentioned that there is no strict division between nouns and verbs in most of the focus languages. Some morphemes such as personal names (e.g. John) are unlikely to occur as verbs. However, personal names can

be used as verbs meaning 'to call someone X'.

Wolff (1973) has a reconstruction of PFL which we will use. His reconstruction is as follows (using our terms):

	Non past	Past	Future	General	Imperative	Subjunctive
AF	-um-	-inum-				
PF	-ən-	-in-	r	-ən	-a	
LF	-an	-in- -an	r	-an	-i	-ay
IF	si-	s-in-i			-an	

r = reduplication of first CV of base.

We have made one alteration to Wolff's forms. Wolff has i not si for the IF prefix. Taiwanese evidence clearly shows that it was *si. This should be reflected as hi in the Philippines, however apart from Tausug hipag- and Waray mahi- (see Zorc 1977:134), it is always reflexed as i. However the Tausug and Waray evidence clearly require us to reconstruct *hi for PPh, as a reduction *hi → *i is one of the most natural sporadic sound changes in the world.

The absence of *-ən in the perfective PF calls for some comment. This absence of *-ən is usually explained as being due to 'zeroing-out' whereby the presence of *-in- caused the phonologically similar suffix *-ən to disappear. However, this explanation is totally inadequate, for if *-in- causes *-ən to disappear, it should also cause *-an the LF marker to disappear. There is a much better alternative explanation. The PF construction is an ergative construction. The connection between ergativity and perfectivity is well known. In Tongan ergative clauses are perfective and accusative clauses are non-perfective (Foley 1976). In some Caucasian languages perfective clauses are ergative and non-perfective clauses are accusative (Dixon 1979). Therefore it is reasonable to suggest that if a perfective clause carries no focus marker for which role is PrP then the unmarked reading of a clause with *-in- only would be that the patient was PrP (i.e. that it was a PF or ergative clause).

So far we have been presenting a picture which views the focus affixes as simply conveying the information that role x is the PrP of the clause. This is a rather simplistic view of the function of the focus affixes. Wolff (1973) shows that in PFL, the LF marker was used for patients of accidental and involuntary actions. The examples below show this:

- SL tingali hibati?an ?a ?atu? pagharampang
 lest hear-LF CM= our conversation
 'Maybe someone will overhear our conversation'
- Ts oana i si nana ahta talui
 no longer TM by her particle at all happen to be heard
 'She never heard it again'

The LF marker was also used to refer to recipients of non-volitional verbs such as forget. The LF marker was used to refer to the thing with reference to which a mutual action takes place, whereas the people engaged in the action were referred to by PF.

- SL pag-?a-?away-an ta ?a nahinabu?
 fut-quarrel-LF CM= what happened
 'We will quarrel about what happened'

SL ʔa-ʔaway-un ku siya
 fut-fight-PF I he
 'I will fight with him'

Wolff states that in many languages the LF marker has extended its range at the expense of the PF marker.

In PFL the instrumental focus marker also covered beneficiaries. It also covered the recipients of verbs of motion away from the Actor (i.e. giving, throwing). The IF marker is also used with verbs such as say, tell, etc. to refer the thing asked, etc. Working on the principle that affixes lose semantic specificity over time, it is probable that these affixes had even more concrete meanings in PFL. Foley (1976) in his analysis of Tagalog shows that in present-day Tagalog one cannot treat the affixes as having only a pure case-marking function. The focus affixes have a discourse function. In most languages, old information occurs clause initially for linkage purposes, but in Focus languages the verb which is new information occurs initially.

The focus affixes, however, refer to the PrP which is in all focus languages usually old information. From the semantics of the affix and the case frame of the verb one is able to tell which NP will be the PrP and so one knows which pieces of information from the previous clause(s) is the centre of this new clause.

Finally in our discussion of the focus affixes we should note that it appears they could all be used as nominalisers in PFL. This fact will prove quite important in our discussion of PAN syntax. Dardjowidjojo (1978), in his discussion of nominal derivation, states that Kə, pəN, pər, -an and combinations of the first three and -an occur as nominalisers. Now -an is obviously a reflex of the LF marker *-an. It forms nouns with the meaning of (i) what one (verb), (ii) things which are (adjective), (iii) place where one (verb), (iv) collection of X. Examples are (i) anjur-an 'suggestion', (ii) manis-an 'sweets', (iii) parkir-an 'parking space', (iv) laut-an 'ocean'. The third use also occurs in Paiwan (Ferrell 1982) and Tagalog (Schachter and Otnes 1972) languages where -an still functions as the LF marker. PəN occurs as focus affix in the Philippines but is not reconstructable for PFL (see section 3.). Kə, on the other hand, is found throughout the focus languages, both alone and in combination with -an. In Indonesian kə- is a frozen form occurring in only three items. In Paiwan its uses are varied (e.g. ka-tiaw 'yesterday', si-ka-taqəd 'reason for sleeping'). In Tagalog, the time words ka-hapon 'yesterday' and ka-gabi 'tomorrow' take it as they do in Paiwan. Tagalog also has forms such as ka-kwarto 'room mate' (kwarto 'room'), karelihyon 'co-religionist' (relihyon 'religion'), katabi 'someone/thing beside' (tabi 'side'). In Palauan klə, from *k-in-a with the perfective infix, forms abstract nouns from stative verbs (Joseph 1975). It also functions as a verbal affix in Paiwan and Tagalog. In Paiwan it functions as an inchoativiser for some stems (Ferrell 1982) and in Tagalog it occurs alone or in combination as Maka-. Alone it means involuntarily doing something; maka also indicates this but it also indicates the ability to do the action of the verb. The meanings attached to this prefix are rather diverse but basically it seems to have two meanings. These are 'state' and 'unmonitored, uncontrolled change of state'.

In Paiwan, Tagalog and Indonesian we also find *ka- ... -an as a nominaliser. This form does seem to have a reconstructible semantic value. In Paiwan Ferrell says that its meaning is 'principal or main manifestation of'. In Tagalog Schachter and Otnes define it to mean 'the class or qualities or group of people/things instanced by what the base designates'. In Indonesian Dardjowidjojo

gives its meanings as (i) product of (ii) state or quality of being (iii) location (iv) collection (v) abstraction. From these descriptions one can see that there is a common idea of 'group or individual which manifests the thing instanced by base'.

In Tagalog and Paiwan *ən can also be a nominaliser. Examples are Pw kən-ən 'food' (kan 'eat'), Tg awit-in 'song' (awit 'sing'). It is used as a nominaliser in Palauan. In Paiwan the AF marker -m- and the IF marker si- can be used as nominalisers, e.g. q/m/uṭuts 'wrestler': quṭuts 'wrestle', si-qunu 'knife': qunu 'cut'. The AF and IF markers can also be used as nominalisers in Tagalog. (e.g. ang pumasok 'the one who entered', ang ibibili niya 'the thing he is buying'). The perfective infix *-in- also occurs as a nominaliser throughout the Philippines.

Therefore it appears that in PFL the focus affixes were bound morphemes which had a definite semantic value which conveyed information about the NPs associated with the verb. They could also be used to form nominals from verb stems. The semantics of these nominals was probably determined by the semantics of the focus affixes.

Having reconstructed these affixes and fitted them into the focus system, it is of great significance for any reconstruction of PAN syntax that we find the same affixes functioning as nominalisers in languages with an Oceanic syntax. Pawley and Reid (1976) point out that *-an, *si- and *-in- are productive and reflected in many Oceanic languages as nominalising affixes and *-ən has traces in the Oceanic languages. There are also traces of *m- in Oceanic (there is replacement of initial p or b by m sporadically – in many focus languages when a stem with initial p or b occurs in AF, the p or b becomes m). The semantics of the Oceanic affixes are relatable to the semantics of the affixes in the focus languages.

The next major fact which must be accounted for in any reconstruction of PAN syntax is the presence of reflexes of the Oceanic transitive suffixes *i and *aki(ni) in Indonesian. In Indonesia these suffixes indicate forms *i and *akən. Pawley and Reid (1976:60, 1976:70 note 15) provide strong arguments that *akən is the original form of the suffix. These suffixes have uses which match those of Oceanic languages. However, languages such as Toba Batak and Wolio which have these suffixes make a three-way distinction against a two-way one in Oceanic. In Toba Batak those roles which are marked by Patient Focus in the focus languages occur with a $\emptyset$ marked verb (no suffix), those which are markedly LF, are marked by i in TB and those which are marked by IF are marked by -hon. In Oceanic *Ci covers both TB i and $\emptyset$.

We may assume that the TB represents the original situation. It is more likely that Oceanic would have extended the range of *i than that TB and to an extent Wolio would have restricted it. The theory which best explains the presence of the focus affixes in the Oceanic languages and the presence of the Oceanic suffixes *i and *akən in Indonesia is Pawley and Reid's (1976). They propose that PAN had clause structures with both Oceanic and focus features, a situation which occurs in languages such as Toba Batak. The reconstruction, which we propose, differs in details from that of Pawley and Reid.

We propose the following clause structures:

1. Stative clauses.

ma-Verb NP*

2. Active clauses.

a) AF clauses

V-um-erb NP_{Pat} ACTOR NP_{Loc} NP_{Inst}
 V-um-erb-i NP_{Loc} ACTOR NP_{Inst} (NP_{Pat}?)
 V-um-erb-akən NP_{Inst} ACTOR NP_{Loc} (NP_{Pat}?)

The perfective has -inum-.

b) PF clauses

Verb-ən NP_{Pat} ACTOR NP_{Loc} NP_{Inst} PP* (non-perfective PF)
 V-in-erb NP_{Pat} ACTOR NP_{Loc} NP_{Inst} PP* (perfective PF)
 Verb-an NP_{Loc} ACTOR NP_{Pat} NP_{Inst} PP* (non-perfective LF)
 V-in-erb-an NP_{Loc} ACTOR NP_{Pat} NP_{Inst} PP* (perfective LF)
 si-Verb NP_{Inst} ACTOR NP_{Pat} NP_{Loc} PP* (non-perfective IF)
 s-in-i-Verb NP_{Inst} ACTOR NP_{Pat} NP_{Loc} PP* (perfective IF)

The development of the Oceanic type syntax would follow the lines suggested by Pawley and Reid. PAN would presumably have had a verb-initial clause structure with free order among the NPs. As has been already discussed the function of focus affixes is to provide old information for discourse cohesion purposes when each clause starts with new information. The basic innovation in Oceanic was the movement of the Actor to the first position in this clause. It seems that the best explanation for this is either that PAN had an alternative construction for sentence-level topics of the form Topic Verb (NP*) with a pronominal trace of the Topic in the clause as occurs in many languages throughout the world, or that some high-order subgroup (POC?) innovated this. It seems most likely that it existed in PAN as it appears that all languages have sentence-level topicalisation procedures and one leaving a pronominal copy in the clause appears to be the commonest cross-linguistically. One of the main functions of topicalised constructions is to introduce new themes in discourse. From this start it is relatively simple for it to spread to introducing themes in general. Once this NP initial construction became an unmarked construction there would be little reason to maintain the focus system, because there would be an alternative system with old information clause initially. Parallel to, or following, this would have a restriction in the roles which could occur in this clause-initial position until only Actors could occur there. This is an unsurprising development considering that Actors tend to act as themes for discourse in the majority of the world's languages. Once these two things had happened we would have clauses with the structures:

Actor Verb $\emptyset$ Patient/Goal X
 Actor Verb i Locative X
 Actor Verb-akən Instrumental X.

To obtain the Oceanic clause structure we only need to collapse the $\emptyset$ and *i categories. This can simply be regarded as example of the spread of locative markers into the range of patients and goals. The non-Actor focus constructions have been lost in Oceanic. The focus affixes have been retained in their nominalising function. The reasons for the loss of the non-AF constructions are not entirely clear. As we have pointed out the development of noun-initial word order would make the focus affixes obsolete to a large extent. The AF marker

m- would have no function because the Actor would always occur initially, and so it is lost almost without trace in Oceanic. When the word order in Oceanic became Actor initial we would have two conflicting systems of discourse linkage with the non-AF clauses. We would have one system which keyed on Actors and another on the focus affixes. Apparently in Oceanic the system of linkage keying on Actors won out and the system of linkage on focus affixes was lost and so the non-AF clause types were lost.

There is one further thing which we must explain in our derivation of POC clause structure. The POC Actor concord pronouns correspond to the PAN PrP pronouns and the Object concord pronouns correspond to the PAN genitive pronouns which covered genitives and non-focussed Actors. As the Actor was PrP in the AF constructions, the fact that the Actor concord pronouns are the old PrP pronouns is to be expected. However, as we shall see, the genitive pronouns did not cover patients in PAN. In some focus languages, though, such as Tagalog, non-focussed actors, patients and genitives form a category (Schachter and Otnes 1972). If this was the case in POC then the use of these genitive pronouns as object concord pronouns in POC is to be expected because the patients would have been non-focussed.

This is how we would derive the Oceanic active transitive clause system described earlier. The intransitive clauses would develop simply from the PAN monovalent clauses by the topicalisation procedure. In this there could be no restriction of initial position to Actors only because some verbs (the process verbs) by their semantics require patients.

This course of development follows the course proposed by Givón in his article "Topic, pronoun and grammatical agreement" (1975), where he says that "when a language reanalyses the topic constituent as the normal subject or object of the neutral, non-topicalized pattern, it perforce also has reanalysed subject topic agreement as subject agreement and object topic agreement as object agreement". We are saying that this is what happened in the development from PAN to POC.

A development similar to this appears to be going on in literary and Manila Tagalog today. In her discussion of Tagalog, Naylor (1975) says that verbs with the IF marker are not used. She also states that Benefactive focus forms were much rarer than she expected and that she could find only four instances of Locative focus constructions in all her texts. The rarity of these constructions in written texts may be related to the extensive use of topicalised 'ay' constructions which have initial NPs, in written texts. The use of ay constructions has not penetrated into ordinary colloquial speech, and in Manila Tagalog the disuse of the focus constructions mentioned above occurs and so there is apparently no necessary connection between loss of specific focus types and the rise of topicalised constructions. The factor which argues that there was such a connection in the development of Oceanic syntax are the obligatory pronominal copies which are as Givón argues concomitants of topicalised constructions.

The development of all the focus languages is relatively simple. The focus languages have simply made AF constructions have the same options as the other types of focus. This gets us to the situation we find in most of the focus languages.

Taiwanese Paradigms (not necessarily complete).

Kanakanabu (Tsuchida 1976)

Independent		Enclitic		
		Nominative	Genitive	Oblique
sg 1	i:ku	-(m)aku	-(m)aku	-ʔikua
2	i:kasu	-kasu	-(mu)su	-kasua
pl li	i:kita	-kita	-(mi)ta	-kitana
le	i:kimi	-kimi	-mia	-kimia
2	i:kamu	-kamu	-mu	-kamua

(mV) occurs after vowels

Saaroa (Tsuchida 1976)

Independent		Enclitic		Oblique
		Nominative	Genitive	
sg 1	iʔaku	-aku	-ku	na + independent forms
2	iʔaa	-u	-u	
pl li	iʔaʔamu	-amu	-ʔamu	
le	iʔaʔa	-ita	-ta	
2	iʔamu	-mu	-mu	

Tsou (Tsuchida 1976)

Independent		Enclitic	
		Nominative + Genitive	
sg 1	aʔo	-ʔoʷʔu	
2	su:	-koʷsu	
pl li	aʔmi	-mza	
le	aʔto	-to	
2	mu:	-mu	

Thao (Li 1978)

		Nominative	Non-Focussed NP	Genitive (Possessor only)
sg 1	ya:kuʔ	ya:kuʔ	ya:kin	na:k
2	ʔi:huʔ	ʔi:huʔ	ʔu:hun	mi:huʔ
3	θi:θuʔ	θi:θuʔ	θi:θuʔ	θi:θuʔ
pl li	ʔi:taʔ	ʔi:taʔ	ʔu:tan	mi:taʔ
le	ya:min	ya:min	ʔa:min	na:m
2	mani:yun	mani:yun	ʔi:human	mani:yun
3	θayθuy	θayθuy	θayθuy	θayθuy

Kavalan (Li 1978)

	Nominative	Genitive
sg 1	-iku?	-ku?
2	-isu?	-su?
3	∅	-na?
pl li	-ita?	-ta?
le	-imi?	-nyaq
2	-imu?	-numi?
3	∅	na qanyau?

Pazeh (Li 1978)

	Nominative	Genitive	Oblique/Topic
sg 1	-aku?	naki?	yaku?
2	-siw	nisiw	?isiw
3	-misiw	nimisiw	?imisiw
pl li	-ta?	nita?	?ita?
le	-ami?	niami?	yami?
2	-mu?	nimu?	?imu?
3	-amisiw	na:misiw	ya:misiw

Saisyat (Li 1978)

	Nominative	Genitive	Oblique
sg 1	yako ^v yao	ma'an	yakin
2	ʃoʔoʋʃoʔon	ni-ʃoʔ	?i-ʃoʔon
3	sia	ni sia	hi sia
pl li	?ita?	mita?	?ini-mita?
le	yami?	ni-aʔom	?ini-aʔom
2	moyo	ni-mon	?ini-mon
3	la-sia	na-sia	hi la-sia

Nataoran-Amis (Teresa M. Chen, personal communication)

	Topic	Nominative	Genitive	Locative	Benefactive
sg 1	ako	kako	nomako (naku)	itakoan	saka + Locative
2	iso	kisu	nomiso (niso)	itisoan	
3	cira	cira	nira	iciraan	
pl li	ami	kami	noniam	itamian	Locative
le	ita	kita	nomita (nita)	ititaan	
2	amo	kamo	namo	itamoan	
3	oqni	koqni	noqni	itoqnian	

Paiwan (Ferrell 1982)

	Nominative	Genitive	Category 3
sg 1	-akən, ti-akən	ku-, ni-akən	t'anu-akən
2	-sun, ti-sun	-sun-, ni-sun	t'anu-sun
3	ti-mad'u	ni-mad'u	t'ai-mad'u
pl li	-it'ən, ti-t'ən	t'a, ni-t'ən	t'anu-it'ən
le	-amən, ti-amən	nia-, ni-amən	t'anu-amən
2	-mun, ti-mun	nu-, ni-mun	t'anu-mun
3	tia-mad'u	nia-mad'u	t'aia-mad'u

Rukai (Li 1975)

	Independent		Dependent	
	Subject	Object	Subject Marker	Personal Possessive
sg 1	ko-n-ako	n-akw-a	-(a)ko, -n-ako, naw-	-li
2	ko-so	mo-sw-a	-so	-so
3 def	(ko-a-ni)	(i-ny-a)	∅	-i-ni
3 indef	(ko-a-da)	(i-da)	∅	-i-da
pl li	ko-ta	mi-ta	-ta	-ta, -i-ta, -mi-ta
le	ko-nai	nay-a	-nai	-nai
2	ko-nomi	nomy-a	-nomi	-nomi
3 def	(ko-l-i-ni)	(l-i-ny-a)	∅	-l-i-ni
3 indef	(ko-l-i-ḍa)	(l-i-ḍa)	∅	-l-i-ḍa

(mon 2pl Li 1975:257 archaic form).

Atayal (Egerod 1966)

	Nominative	Genitive	Locative	i forms
sg 1	saku? ku?	maku? mu	knan	kuziŋʷkun
2	su?	su?	(i)sunan	isu?
3	hia?	nia?	hian	hia?
pl li	ta?	ta?	?tan itan	ita?
le	sami	mian	sminan	sami
2	simu	mamu	smunan	simu
3	hga?	nha?	hgan	hga?

Sediq (Asai) [This paradigm is not accurately or fully recorded]

	Nominative	Genitive	Locative
sg 1	jakuzku	nakuvmu	kenan
2	isuvsu	nisuvsu	sunan
3	hiḍa	niḍa	
pl li	yami	mian	
le	ita	nitaṽta?	
2	yamu	namu	
3	dahiḍa	nuhiḍa	

Bunun (Jeng 1977)

	Nominative	Oblique + Contrastive	Accusative enclitic
sg 1	sak ^h ?a ^h ak	ŋaku	-ku
2	?asu ^h ? ^h su?	su ^h u	-su
pl 1i	?ata?	?ita?	-ta?
1e	sam ^h ?a ^h am	ŋami?	
2	?amu?	mu ^h u	-mu
sg 3	sia	?ista	-ta?
	siata?	?istun	
	?aipa? (far)		
	?aipi? (near)		
	?aipu? (far)		
	?aipun		
pl 3	nai	?inta	
	naip	?intun	
	naita?		
	?iŋka?		
	?iŋkun		

Genitive

	pre-nominal	post-nominal
sg 1	inak	nak
2	isu ^h u	su ^h u ^h su
3	isiata?	?itu?
pl 1i	imita?	mita?
1e	inam	nam
2	imu ^h u	mu ^h u ^h mu
3	inaita?	naita?

2.2. PAN pronouns

We will now go on to consider the pronominal system which can be reconstructed for PAN. As the system we have reconstructed for syntax is basically that of a focus language, we would expect that the pronouns would have categories similar to those found in the focus languages. The focus languages have essentially three categories for pronouns. The categories are widely called nominative, genitive and oblique. The nominative form is used when the pronoun is in focus. The genitive covers non-focussed actors as well as genitives, while the oblique covers all other pronoun functions including absolute genitives 'mine, yours' etc. The PAN pronouns which we reconstruct for these categories are not necessarily reflected in these categories nor are the categories themselves necessarily reflected. We will reconstruct four sets of pronouns; one for the nominatives, two for the genitives, one a bound set and the other free, and we will reconstruct a fourth set which may or may not have covered all the range of the third category.

The fullest discussion of the PAN pronouns to date is Blust 1977. He proposes the following paradigm:

	nominative	genitive
sg 1	i-aku	i-ku/ni-ku
2	i-su	i-su/ni-su
3	t'i-ia	i-a/ni-a
pl 1i	i-kita	i-ta/ni-ta
1e	i-kami	i-mi/ni-mi
2	i-kamu	i-mu/ni-mu
3	t'i-ida	i-da/ni-da

We will examine the forms of the genitive pronouns first as these appear to have undergone the least alternations since PAN. Also evidence from the genitive paradigm will be important in determining the nominative paradigm precisely because the genitives have undergone less change. Blust's genitives correspond to our free form genitives. Blust's genitives in *ni- are reflected in many languages such as Paze and Thao. Blust's reasons for reconstructing the *i genitives are that i occurs as genitive article, as well as ni, in Raluana, Bugotu (Oceanic languages), various Batak languages (Sumatra) and Tontemboan (Sulawesi). He finds support for it in alternations between nX genitives following vowels and X genitives after consonants in Javanese, Proto-Ambon, Buginese and Toba Batak.

However, the only form Blust cites which must descend from an *i genitive is Javanese e (from *ia) postconsonantal (ne postvocalic) 3rd sg genitive. The other alternations kuŋku (1st gen) and taŋta (1pl incl) provide no evidence for an i genitive. Furthermore, as we will see in our discussion of the nominative forms, it is doubtful whether PAN had third person pronouns. In any case the existence of postvocalic nV and postconsonantal V variants of suffixed genitive pronouns in one language is not a firm basis for positing that PAN had post-vocalic ni and postconsonantal i genitives. The Javanese e form could have arisen from a reduction of Cne to Ce. However, Reid (1979) states that there is evidence from Bontok to support Blust's -i genitives. Therefore for the moment we will regard the status of Blust's i genitives as uncertain.

On the other hand the genitives reconstructed by Dahl (1976) and Dyen (1974) are reflected in many languages. The forms reconstructed are:

sg 1	ku
2	su
3	n'a
pl 1i	ta
1e	mi
2	mu
3	da

Forms such as these occur in Tsou, Saaroa, Javanese and Magori and many other languages. These languages do not necessarily have all these pronouns but on the basis of many languages the above pronouns can be reconstructed.

Blust makes a number of points about this paradigm. The first is that *n'a is better reconstructed as *nia, a form which is segmentable into *ni the well known genitive marker and *a an element which occurs in the third person nominative pronouns. The second is that outside Taiwan the 2nd pl clitic mu is frequently reflected as a 2nd sg. As Blust points out, this is probably due to a politeness shift, a well known cross-linguistic phenomenon. The third point which we will note is that there is no evidence in Taiwan for a 3rd pl genitive *-da and therefore we will not reconstruct it for PAN.

A number of Taiwanese languages provide evidence for another genitive. Saisyat maʔan, Atayal makuʔ and Kanakanabu (m)aku (m after vowels) provide evidence for a 1st sg genitive in m. However, these three forms are not relatable. While Atayal and Kanakanabu point to *maku, Saisyat cannot. In Kanakanabu the 2nd sg and 1pl incl, (mu)su and (mi)ta have m alternatives following vowels. Saisyat and Bunun also have mitaʔ 1pl incl gen and Thao has mi:huʔ 2sg gen. These forms all argue for some sort of m genitive. However, its form and functions are now irretrievably lost. It may have been the original pre-PAN genitive now left only in traces.

The form and function of the ni pronouns is on the other hand recoverable. While we agree with Blust that there were ni pronouns, we do not agree with him as to their form or functions. Pazeh nakiʔ, Thao na:k, Bunun (i)nak, (1sg gen) Pazeh niamiʔ, Thao naim, Bunan (i)nam, Saisyat ni-aʔam (1pl excl) all point towards forms *ni-aku (1st gen) and *ni-ami (1pl excl gen) consisting of *ni the genitive marker and a nominative root. As the 1pl incl nominative root corresponding to *aku and *ami is *ita we will get *nita and we cannot say from this form whether it is *ni-ita or *ni-ta. However, *ni-ami and *ni-aku point towards *ni-ita unequivocally. As the second person nominative roots are *su and *mu we will get *ni-su and *ni-mu anyway. Therefore we reconstruct the two genitive pronoun paradigms as:

sg 1	ku	ni-aku
2	su	ni-su
3	n̄ia	ni-a
pl 1i	ta	ni-ita
le	mi	ni-ami
2	mu	ni-mu

The question naturally arises as to what the differences between the two sets was. The difference, as we have indicated, is that the ni set are free-form pronouns and the other set are bound. Dyen (1974) reconstructed the first set as clitics and it can be seen that they are bound forms in the Formosan languages where they occur. On the other hand, the ni forms when they occur in Pazeh, Thao and Saisyat are free forms. The Bunun i genitives are not related to Blust's proposed *i genitives. They differ in form and function. Jeng (1977) says that they occur after the attributive particle tuʔ and preminally with the consonant forms occurring postnominally.

We will now go on to discuss the pronouns for the third category. These pronouns are reflected in many languages such as Paiwan, Pazeh, Ivatan, Tagalog, Okolod Murut, Cebuano and Malagasy (see Blust 1977). The forms that can be reconstructed are:

sg 1	akən
2	sun
pl 1i	itən
le	amən
2	mun

These forms are composed of a corresponding nominative root and the PF suffix *-ən (Blust 1977). The combinations are:

sg 1	aku + ən
2	su + ən
pl 1i	ita + ən
le	ami + ən
2	mu + ən

The difference in results of adding *ən* to the first and second person forms is probably best explained by a stress difference. The first person forms are disyllables and the second person forms are monosyllables. If the disyllables were initially stressed then the difference is explicable.

These forms have many functions in the various languages where they occur. In Tagalog they are the oblique forms; in Paiwan the base; in Okolod Murut the goal focus forms. However, in all languages where they occur, either the base alone or some affixed form of this base (whatever other bases occur) covers the meaning of the absolute genitives 'mine, ours', etc. Therefore this function must as Blust points out be attributed to these forms in PAN. This function follows naturally from the semantics of a combination of the root and the PF marker **ən*. Cross-linguistically the connection between 'It is to me' and 'It is mine' is well known. This pronoun form presumably also is an oblique form at least for patients and goals. Therefore the use of this form as an absolute genitive is unsurprising. We do not know if this form covered all the range of the third category in PAN. This combination of a pronoun root and a focus affix may seem odd but in Atayal and Sediq there are forms of the pronoun which consist of a root and the LF marker *-an* (Egerod 1966). There are similar forms in Amis. These forms are the oblique forms. In Thao and Saisyat these **ən* forms function as obliques for the 1st sg and in Thao for the 2nd sg as well. Furthermore, as we have already seen, though there are verb and noun slots, the fillers are not so categorised normally and it appears that in PAN 'verbal' and 'nominal' affixes (the PF marker *-ən* and as we shall see the noun focus particle *i*) could be affixed to the pronouns. In Paiwan these forms have completely replaced the original nominatives and are now bases for a restructured paradigm. Reflexes of **akən* as the first person nominative are found in many languages. In Tagalog *amin* and *atin* (from **amən* and **itən*) are the non-nominative bases.

It appears that there is a tendency for these forms to extend their range. One possible explanation might be the use of these forms in cleft-like constructions such as 'it was me who did it' where the oblique form is the focussed form of the lower clause. This usage could explain how these oblique forms spread.

We will now go on to consider the forms of the nominative pronouns which Blust reconstructed as:

sg 1	i-aku	
2	i-su	i-kasu (polite)
3	t'i-ia	
pl 1i	i-kita	
le	i-kami	
2	i-kamu	
3	t'i-ida	

We have left the discussion of these forms to last because they show the most variation and are therefore the most difficult to reconstruct. Also paradigmatic evidence from the three paradigms we have already reconstructed will be important. We will consider the forms of the pronouns separately. The form of the first singular is not subject to much dispute. The form reconstructed by Brandstetter and Dempwolff was **aku*, reflexes of which are found in many languages such as Tsou, Javanese, Malay, Tagalog, Fijian and Samoan. A form **i-aku* with the prefixed common noun focus marker **i* is found in many languages such as Thao, Saisyat, Bareʔe, Bugis and Jamedna (Blust 1977). Kavalan and Kanakanabu *ikuʔ* and *i:ku* are presumably further developments from this. Atayal *sakuʔ* corresponds to Sediq *yaku* and Atayal *sami 1pl excl* and *simu 2pl*

correspond to Sediq yami and yamu. As Sediq and Atayal form a subgroup (see p.89) it appears that one should reconstruct Proto-Atayalic *yaku. The same explanation may also hold for the Bunun forms sak^v?a^oak. In Bunun $y \rightarrow \delta$. These forms could also represent *t'i-aku with the proper noun focus marker prefixed. This form should give siak in Bunun but a change siak $\rightarrow$ sak is not particularly unlikely. In Ilokano a form siak, siaku-n occurs which could also derive from *t'i-aku. However, it is also possible that it derives from Proto-Cordillerean *si-yakən (Reid, personal communication). There does not appear to be enough evidence to posit the existence of a PAN *t'i-aku, as sak in Bunun could possibly arise from *yak as it does in Atayal.

The second person is the most problematic one in AN pronominal reconstruction. The reflexes indicate three different roots *ka, *i-su and *mu which combine in many different ways and appear to swap over in number in the different languages. We will consider the singular forms first. *ka is found in Tsou and many Philippine languages. *i-su is found in many Formosan and Philippine languages and in Malagasy. This appears to consist of the focus marker *i and a form *su which occurs in Tsou, Bunun and Atayal. A combination of the two *kasu is reflected in many non-Formosan languages. A form *ikasu is reflected in Kanakanabu, Tagalog, Fijian and other languages. This means we have forms *ka, *su, *i-su, *kasu, *i-kasu for the 2nd sg. Blust proposes that *(i)-kasu was the polite pronoun in PAN with *(i)-su being the familiar one. There is, however, no language which reflects *(i)-kasu as a polite pronoun and this explanation does not tell us what the functions of *ka and *(i)-su were. In Tsou ko (*ka) and su (*su) occur in morphophonologically conditioned environments; su occurs after a focus particle ending in o or oh and ko elsewhere. However, it would be exceedingly dangerous to say that this reflects the PAN usage. Finally it must be mentioned that Ivatan has a singular pronoun ?imo which corresponds to a plural elsewhere. It is generally assumed that the Ivatan use arose as a result of a politeness factor. Tagalog reflects both *i-kasu and *ka without apparently there being any particular difference in politeness.

I think that one must also consider the plural forms *kita, *kami and *kamu. Dahl (1976) has pointed out that the Formosan languages appear to reflect forms without the k and that the k may reflect some sort of old pronominal prefix. If we examine the Formosan languages we find, corresponding to *kita, Bunun ?ata? and ?ita?, Kavalan ita?, Thao ?i:ta?, Saisyat ?ita?, Paiwan it'en, Atayal ita?, Saaroa i^oata, Rukai ko-ta and Tsou a^oto. These forms point towards *ita but Tsou, Rukai and Bunun also indicate *ata (in Tsou initial k is lost so the Tsou form could reflect *kata). If we examine the reflexes of the 1pl excl in Formosa we get Bunun sam^v?a^oam, Kavalan imi?, Thao ya:min, Saisyat yami?, Paiwan amən, Atayal sami, Pazeh yami?. These forms all point to *ami. Finally if we examine the forms for the 2nd pl we find Bunun ?amu?, Archaic Rukai mun, Paiwan mun, Atayal simu, Pazeh ?imu?, Saaroa i^oamu, Tsou mu:, Kavalan imu?, Saisyat muyu. These forms indicate a form *mu (the i which is found here and on the Formosan forms for kita and kami is of course the noun focus marker). It is not only the Formosan languages which lack the k. Ivatan (Reid 1971) has imo (2nd sg - corresponding to 2nd pl in Formosa), Subanon has yami, yamu and ita.

The parallelism between *su: *kasu and *mu: *kamu is too obvious to miss. The similar parallelism *ita: *kita and *ami: *kami is also obviously the same phenomenon. It appears therefore that an element *ka which is found as a 2nd sg pronoun could be prefixed to the other 2nd sg *su, the 2nd pl *mu, the 1pl incl *ita and 1pl excl *ami. The common factor for the prefixing of *ka seems to be

the 'second person', with *kami being due to paradigmatic pressure from *kita and *kamu.

There is, however, further data which must be brought into consideration when dealing with the question of *ka and the second person. Saisyat moyo, Itbayatan imi:yu, Sambal muyu, Casiguran Dumagat moy and Agta moy (Reid 1971) all point toward a 2nd pl pronoun *muyu. Ivatan has a form ?inio?, Binuhid has sinyu and Tagalog has a benefactive form ?inyo which point to a form *inyu. Now reflexes of a form *kamuyu are found in AN languages such as Agta. Reid (1974) has reconstructed a form *kamuyu for the 2nd pl in Proto-Cordilleran (a Philippine subgroup). A form kayu occurs in a number of Cordilleran languages as well as in the neighbouring Tagalog and Kapampangan. Harmon (1974) in her reconstruction of Proto-Manobo (another Philippine subgroup) has an emphatic form *kiyu which is presumably related to *kayu. Finally in Tagabili the 2nd pl pronoun is yuh (Reid 1971). All of these facts suggest that there was a 2nd pl pronoun or pronoun suffix *yu.

All of these facts combine to make determination of the forms of the second person pronouns in PAN rather problematic. For the 2nd sg we have *ka, *su and *kasu; for the 2nd pl we have *mu, *yu, *muyu, *kamu, *kayu, and *kamuyu. If we accept that the *ka forms of these pronouns arose as a result of prefixing *ka 2nd sg to stems, then we must regard the non *ka forms as older. Support for this is provided by the fact that in the paradigms we have already reconstructed there is no trace of *ka.

This leaves us with *ka and *su for the 2nd sg and *mu, ?*yu and *muyu for the 2nd pl. This *yu form is rather problematic. Outside Formosa *s is always reflected as either h or ø and so *i-su gives i(h)u. Now Tagalog has a form iyo (*iu) which is the locative form of the 2nd sg and Malagasy has a form iu which was a 2nd sg pronoun. The question arises as to whether this iu from *i-su is the same as the *yu we have reconstructed as occurring in the 2nd pl. Blust (1977) argues that they are, but as he points out we cannot explain the Saisyat form moyo 2nd pl this way. In Saisyat *su gives [oʔon (s → j), u → o] Formosan languages frequently add ? initially and finally, the -on is probably *-an. However, elsewhere in the paradigm we have [oʔ). This, I think, forces us to conclude that the *yu found in the 2nd pl is not related to the *iu 2nd sg pronoun found outside Formosa. These are two cases where *yu occurs alone as a pronoun. The first, which we have already mentioned, as Tagabili yuh. Now this could be taken as good evidence that *yu occurred as a free-form pronoun. However, the Tagabili (Reid 1971) 1pl excl is mih which must derive from *ami or *kami. This raises the possibility that Tagabili *yuh is from *kayu or some other form ending in yu. Therefore Tagabili cannot be said to provide good evidence for an independent *yu. The second situation is in certain Manobo languages (Harmon 1974) where it occurs as a clitic pronoun. As clitics are usually reduced forms of pronouns, it appears best to say that there is not enough evidence that *yu occurred as an independent 2nd pl pronoun. This leaves us with *mu and *muyu. Of these *mu is older both because it is unaffixed and because it occurs in the other paradigms we have reconstructed and *muyu does not.

For much the same reasons we would say that *su is the original PAN 2nd sg not *ka. *su occurs in the other paradigms we have reconstructed and *ka does not and it does not occur as a prefix on other pronouns as *ka does. The general trend in the non-Formosan AN languages for the ka prefix pronouns to replace the non-prefixed ones (in most languages *kita, *kami, *kamu have replaced *ita, *ami and *mu) argues that *su is the original form which was replaced by *kasu

in most languages. (Dempwolff (1934-38) who did not have Formosan data reconstructed *kaw from *kasu). It is an incontrovertible fact that *ka was a 2nd sg pronoun before the break up of PAN. The Tsou and Philippine evidence requires us to claim this.

The exact function of *ka in PAN does not appear to be determinable from the evidence available. The distinction between *ka and *su may have been the familiar one of deference/politeness but no present-day AN language reflects this. As already mentioned, Tsou is the only AN language which still has both as 2nd sg pronouns and in Tsou the conditioning for the use of one or the other is morphophonemic. This may have been the conditioning factor in PAN with *ka occurring after certain particles or in certain constructions and *su elsewhere. This could explain how *ka became a prefix. If it signalled 'second person singular' in certain environments, then it could probably generalise to signal 'second person' in those environments and so be added to *mu and *ita and from there be added to *ami by paradigmatic pressure so it lost its meaning. On the other hand *ka could have been some morpheme which by its own semantics was naturally associated with the second person (like a deictic). It could have been prefixed to the second person pronouns (which we take to include *ita lpl incl) in certain circumstances. As the singular is unmarked in the opposition plural/singular, the occurrence of *ka alone would presumably be interpreted as singular. This could explain the occurrence of *ka, *su and *kasu as opposed to *kita : *ita and *kamu : *mu. We should note that Amis has a series of ka versus non-ka pronouns. In Amis the non-ka forms are topic pronouns whereas the ka forms are nominative. However, it is unlikely that this is a reflection of the PAN differentiation. No other AN language has forms kako (from *ka + *aku) and kiso (from *ka + *i + *su). They are presumably innovations resulting from paradigmatic pressure of non-ka topic and ka nominatives in the first and second plural. This differentiation in the first and second plural must be an Amis innovation otherwise it would have left some traces elsewhere in AN. However, there are probably many other equally plausible or implausible scenarios explaining the relationship between *ka and *su. All we can do on the evidence available is reconstruct both. We must also reconstruct the form *kasu.

This means that we are suggesting the following first and second person nominative pronouns.

- sg 1 aku/i-aku
- 2 su/i-su, ka, kasu/i-kasu
- pl 1i ita, kita/i-kita
- 1e ami/i-ami, kami/i-kami
- 2 mu/i-mu, kamu/i-kamu

There was presumably some difference between the forms with i- and those without i-. *i is the common noun focus marker (i.e. it signals that the NP following it is by nominative). However, these pronouns without the i signal this anyway and so this cannot have been the function of the i pronouns in PAN. Yet their function must have been similar to that of the plain pronouns because they have replaced the plain form pronouns in many languages. Two languages which provide us with evidence as to the nature of the distinction between the *i and the plain pronouns are Atayal and Tagalog. Atayal has two sets of pronouns which Egerod (1966) calls nominative (corresponding to our nominative) and 'nominalised'. The two sets are the same except for the 2nd sg isu (nominalised) vs su (nominative), lpl incl ita? (nominalised) vs ta? (nominative) and 1st sg kuziʔʔkun (nominalised) vs sakuʔʔku? (nominative). The 1st sg forms are not relevant here but the difference between isu and su and ita? and ta? is of great interest. Egerod says that the nominalised forms occur "as complements

or in exposure". From the sentence examples in Egerod (1966) it appears that the nominalised pronouns function as sentence-level topics; reflexives; and contrastively. Much the same thing occurs in Tagalog for the 2nd sg which has two forms *ikaw* (from **ikasu*) and *ka* (from **ka*). *ikaw* functions as a sentence-level topic or contrastively while *ka* is used elsewhere.

There can be little doubt from these facts, I think, that the **i* pronouns must have functioned this way in PAN. The **i* pronouns are not the sort of phenomena which would arise independently in two languages. The functions of the **i* pronouns follow reasonably from the semantics of a combination of the focussed forms of the pronouns and the focus marker **i*. The function of **i* is to tell the hearer that the nominal following it is the centre of attention for the clause. Now if we already have special pronouns forms which convey this information, a combination of the two forms would surely indicate that there was something special about this pronoun as centre of attention for the clause. Contrastiveness is the most obvious example of this. Sentence-level topics and reflexives are also good examples. In each case there is something unusual and all three interrelate. Sentence-level topics and reflexives can be used contrastively: 'I was the one who did it', 'I did it myself'.

From these functions the **i* pronouns could easily spread to take over the functions of the plain forms. The connection between sentence topics and nominatives is very strong. Nominatives are unmarked themes and topics introduce new and contrastive themes. The contrastive use is quite likely to be devalued over time by overuse (consider 'terrible' and 'awful' in English). The use of these forms as sentence-level topics would also explain the forms in Kn and Saaroa. In these languages the *i* form pronouns are free form pronouns, and the plain forms are bound. The *i* form pronouns would be most unlikely to develop into clitics because of their functions whereas the plain forms could easily do so. One thing should be noted about the **i* forms. There is no evidence for a form **i-ka* which is another argument that **ka* is a later development as a 2nd sg and that **su* is the original pronoun.

This leaves us with the third person pronouns to reconstruct. Blust has reconstructed **t'i-ia*. This form is reflected by Atayal *hia?*, Saisyat *siya*, Tondano *sia* and so on in many languages. Many other AN languages such as Malay, Ibanag, Nias and Gayo reflect simply **ia*. Now this pronoun is fairly easily analysable into a combination of the common noun focus marker **i* and the demonstrative **a*. **i* of course, occurs on all the other pronouns. Now this argues that PAN did not really have third person pronouns but had instead deictics which could be used to refer to entities not covered by the first and second person pronouns. The reanalysis of the demonstrative as a pronoun must have taken place in PAN. The reanalysis, however, was not a simple reanalysis of the demonstrative **a*, the form reanalysed was the compound **i* + **a*, a focussed and possibly nominalised form (meaning 'that one').

The question arises as to why the third person is the only person with a variant in **t'i*. Now **t'i* is usually reconstructed as the personal noun focus marker with **i* being reconstructed as the common-noun focus marker (i.e. when these morphemes occur before a noun they indicate whether that noun is common or proper and that it is by nominative). However, there are good arguments for reconstructing **a* as the focus marker for common nouns. It occurs as such in Paiwan, Palauan, Tagalog and other languages. Until now we have been using Reid's (1978) reconstruction of the personal noun FM as **t'i*, the common noun FM as **i* and **a* as a demonstrative. However, if these were the PAN forms why do the first and second person pronouns which are inherently proper take the common-

noun focus marker *i, and there is no evidence with the possible exceptions of Atayal and Bunun that they took *t'i.

Now, it is a well-known fact that pronouns tend to preserve older forms. The fact that *i is prefixed to the first and second person pronouns in combination with the fact that a number of languages reflect *a as the common noun FM makes it seem quite probable that at some stage of PAN *i was the proper noun FM with *a being the common noun FM. However, there are a number of facts which make this analysis problematic. Firstly there is the fact that *t'i can be reconstructed as the proper noun FM in PAN. It is reflected as such in Formosa, in the Philippines, Borneo and Madagascar. However, *i occurs as the proper FM in Gaddang, Umiray-Dumagat and then in Kapampangan which is in a different Philippine subgroup. But it also occurs as the common FM in many Philippine languages and Reid reconstructs it as such for PPh. Secondly Reid reconstructs *a as the PPh ligature (the ligature in AN is a marker which occurs in NPs to show that the head N and its attribute form an NP). It has this function in a number of Formosan languages such as Kavalan and Thao where it is not the common FM. However, in Palauan a is the article and not the ligature. Furthermore, why should one combine *i and *a to form a third person pronoun if they were simply the personal and common versions of the FM. It makes more sense for a combination of *i as a proper FM with *a as a demonstrative (Reid's proposal) to function as a third person pronoun (as we pointed out earlier with a meaning something like 'that one').

There is another area which needs considering before a solution to these problems is proposed. This is, of course, the third person plural. This has been reconstructed as *t'i-da. This form has generally been accepted as the PAN 3rd pl. However, there are reasons to dispute this assignment. The only possible cognate for *t'i-da as a 3rd pl in Taiwan is Atayal hga⁷. The development d → g is otherwise unknown in Atayal or AN. It is phonologically unlikely and so cannot be explained as some natural phonological change occurring sporadically in the pronouns. Therefore we must reject the Atayal form as non-cognate.

The obvious Taiwanese cognate for *t'i-da is the Amis 3rd sg tsira. This immediately raises the question of whether *t'i-da was a singular or plural and if it was a singular what *ia ~ *t'i-a was. There is another fact which is relevant in deciding these questions. This is the lack of third person pronouns in a number of Taiwanese languages. Bunun and Rukai have no third person pronouns, only deictics. Tsuchida (1976) does not give third person pronouns for Tsou, Kananabu and Saaroa. The forms of the third person pronouns in Paiwan, Pazeh, Thao and Kavalan are unrelatable to one another or anything else in AN. The Atayal and Amis third plurals are unrelatable to anything else in AN. In Saisyat and Sediq, *t'i-a is the root for both singular and plural.

There are two possible explanations for this state of affairs. Either the third person pronouns have been lost in many Taiwanese languages or they have been innovated as pronouns since PAN. The evidence clearly favours innovation. Accepting that the third person pronouns are innovations helps us with finding a solution to the problem with *i and *t'i, raised earlier. The best solution to that problem seems to be that *t'i is a PAN innovation. In pre-PAN times there was no proper vs common distinction in the construction markers, *i was the only FM. This explains why it occurs on the pronouns, which are, as we have already pointed out, inherently proper. Some time before the break-up of PAN *t'i was innovated as the proper FM. This naturally led to the creation of a *t'i-a form in competition with the already existing *i-a. We can propose a

plausible origin for *ia, *t'i-a and *t'i-da. We have suggested that *a was a PAN or pre-PAN demonstrative and that the focussed form *ia would be a natural choice for a morpheme to refer to the third person entities. Once the proper marker *t'i was innovated a form *t'i-a would naturally arise in competition with *ia.

The origin of *t'i-da is harder to determine. The *da could come from two sources. One is the da plural marker which occurs in the Philippines: *t'i-da could be simply 'those plural ones'. However, the occurrence of *t'i-da as a singular in Amis causes problems. The development of a 3rd pl into a 3rd sg is not something which is known to have occurred in the development of any other language. Another possibility is that *da was a deictic element like *a. A deictic formative da occurs in Hanunoo; Bisayan has da' which means 'near you' and Paiwan has a deictic tsad'a meaning 'far away'. If these forms are cognate, they could represent developments of a *da deictic. However, without a cognate from a language which preserves the distinction between *d and *d', we cannot be certain that the reflexes in the languages above do not represent a form *d'a. Therefore, it appears that for the moment we will have to leave the origin of *da as uncertain.

The occurrence of *t'i-da in a smaller number of languages than *ia ~ *t'i-a supports the view that *t'i is an innovation. As *t'i-da is the rarest of the third person forms, it must have been innovated later. Its occurrence with *t'i only makes it clear that by the time *da came to be used pronominally, *t'i was the proper case marker.

The non-occurrence of an oblique -ən form for the third person and the non-occurrence of a bound genitive separate from the free *ni- genitive, also provide strong evidence for the innovation of the third person pronouns. The situation we are proposing for pre-PAN is that there was a deictic *a which could be used pronominally in its focussed form. Later *t'i was innovated as a proper case marker and a form *t'i-a started to compete with *ia. Later still, an element *da came to be used pronominally in its *t'i form.

This means that we are proposing the following pronoun system for PAN:

Category 1 (nominative)

sg 1	aku/i-aku
2	su/i-su, ka, kasu/i-kasu
pl 1i	ita, kita/i-kita
le	ami/i-ami, kami/i-kami
2	mu/i-mu, kamu/i-kamu

Category 2 (genitive)

	bound	free
sg 1	ku	ni-aku
2	su	ni-su
pl 1i	ta	ni-ita
le	mi	ni-ami
2	mu	ni-mu

Category 3 (oblique)

sg 1	akən
2	sun
pl 1i	itən
le	amən
2	mun

and forms *ia*, *t'i-a* and *t'i-da* (nominative) *ni-a* and *ni-da* (genitive) which referred to third person humans. It should be noted that *ka* prefixation must precede the innovation of **t'i* because the *ka* pronouns are still prefixed with *i*. This means that *ka* prefixation is very old, which could explain the problems we had in determining its semantics.

2.3. PAN construction markers and deictics

From our discussion so far, we can see that there were at least three nominal categories which must have had separate construction markers (CMs). These are the three categories which occurred in the pronouns. We have already seen that in pre-PAN, the CM for Category 1 was **i*. In later PAN, there were two markers **t'i* for proper nominals and **i* for common nominals. The most widespread CM for the second category is **ni* (found in Paiwan, Tagalog, Toba Batak, Malagasy and Fijian). Blust 1977 also provides evidence for an **i* CM which may have been a postconsonantal variant of *ni*. Reid provides supporting evidence for this from the Philippine languages (Reid 1979).

In the languages where a common vs proper distinction is made, **ni* is the proper CM. The best candidate for the corresponding common CM is **nu*. Reid 1978 provides evidence for both **nu* and a form **na*. Both forms are widely reflected in Austronesian. However, **na* could be regarded as a reduction of **nua* (the form which is found in Paiwan). This form can be analysed as **nu* plus the **a* we have already encountered.

If pre-PAN did not have a common vs proper distinction in the CMs for category 1 then it is most unlikely that it had them for category 2. The difference between the two categories is that the nominative proper CM was innovated whereas the common genitive CM was innovated. This difference is probably due to the fact that Actors and Possessors are usually human and so the undifferentiated genitive marker **ni* would probably become a proper CM when the proper vs common distinction innovated in the nominatives was extended to the genitives.

From our discussion of clause structure it will be obvious that there are at least two CMs in the oblique category. These are of course **i* and **akən*. **i* would have marked non-PrP temporals and locations; **akən* would have marked non-Prp instrumentals, benefactives and comitatives. Another locative marker must be reconstructed for PAN. This is **di* which is reconstructed by Reid (1978) for PPh. It also occurs in Toba Batak and Pazeh. The difference in function between **i* and **di* is unknown. It is not known what CM was used to cover non-focussed patients and goals. The Toba Batak evidence mentioned earlier suggests that they were possibly simply bare NPs. However, it is equally possible that they were covered by the second category CM as occurs in many languages such as Paiwan and Tagalog.

We have already mentioned that **a* occurs as the common FM in Paiwan, Tagalog and Palauan. However, there are a number of other facts about **a* which point to it being a demonstrative, even though it is not reflected as such in any AN language. Firstly, as already pointed out, **a* must have been a deictic for the semantics of **ia* and **t'ia* to make sense. Secondly, Reid (1978) reconstructs *a* as the PPh ligature. A ligature is a morpheme which signals that two items are bound together as an NP. Ligatures occur widely in the AN languages (see Foley 1976). It occurs as a ligature in Pazeh (Li 1978) as well. It occurs as a ligature and the common FM in Paiwan, and Kavalan (Li 1978). In

Paiwan *a also functions as the morpheme which marks that two constituents are in an equational relationship. In Palauan *a is the article but not the ligature. However, there are further complications to this. Blust (1974) has reconstructed a PMP ligature *ŋ(a). Some languages have this and others, such as Ivatan, Maranao, Bilaan and Yogad, have only *a. The only way to explain these developments is to posit *a as a demonstrative originally as Reid (1978) has done. The use of demonstratives as ligatures is well known in many languages of the world ('that man', 'the man that I saw' English relative pronouns are equivalent to AN ligatures). The development of demonstratives into articles is also equally well known.

The exact function of *a in PAN is uncertain. It was clearly a demonstrative (probably a distal demonstrative in pre-PAN). Reid (1978) suggests that *a was the PAN ligature and that the a articles found in Paiwan, Tagalog and Palauan arise from reductions of *na (the PAN distal demonstrative). This hypothesis seems to involve unnecessary complications when there is already a more acceptable source for these forms. It seems better to accept *a as the PAN ligature and a demonstrative.

There are two other demonstratives we can reconstruct for PAN. One, already mentioned, is the *na distal demonstrative reconstructed by Reid (1978). The other is *tsu, which occurs as both a proximal and distal demonstrative in various AN languages (tsu 'this' Paiwan, itu 'that' Indonesian). However, as *na is nearly always reflected as distal, it is perhaps better to reconstruct *tsu as proximal.

Therefore we have a PAN CM system

	common	personal
FM	i	t'i
Gen	nu	ni
Loc		i, di
Inst		aken

and two demonstratives *tsu and *na and a ligature/demonstrative *a.

3. SUBGROUPING THE AUSTRONESIAN LANGUAGES

3.1. The PMP subgroup

We will now consider whether there is enough evidence for us to say that the MP languages form a subgroup. We will consider the phonological and grammatical evidence separately, taking the phonological evidence first.

3.1.1. Phonological evidence

The MP languages all show the following mergers:

- *t and *ts merge to *t;
- *l and initial *ʔ to *l;
- *n and non-initial *ʔ to *n.

Change a) is not particularly convincing as it is the sort of change which could arise in a number of different languages independently. However, changes b) and c) which are a split merger of *ʔ, *l and *n to provide convincing

evidence for a PMP subgroup. A split merger is not the sort of thing which arises independently as a result of drift. Some languages occasionally show no reflex for initial *ʔ sporadically. The examples Dahl cites are *ʔaŋuy 'to swim', *ʔu(n)tuŋ 'mortar', *ʔabuq 'to anchor' and *ʔa(n)tak, *ʔitan 'to use a pestle' which show n reflexes in some of the languages of Sulawesi, the Philippines and Chamorro. However in three cases, ʔaŋuy, ʔu(n)tuŋ and ʔa(n)tak, the ʔ is usually followed by a nasal (in most languages the n in (n) is present). Furthermore sporadic alternations between n and l (which is the reflex of initial *ʔ outside Formosa) occur elsewhere in AN. TB has lanok 'fly' from *lalæg 'fly' and Tg has 'a-ninao' 'explain' from *linaw 'realise' (Dempwolff 1934-38). These five words also exhibit the normal l reflexes in many other MP languages such as Tagalog, Javanese and Malagasy. Therefore I do not think that these five words can be said to present a challenge to the split merger.

Another equally, if not more, convincing piece of evidence for a PMP subgroup is the innovation of *n'. This phoneme must be reconstructed for the non-Formosan languages, but as we have demonstrated it appears to be an innovation of these languages. As we have already mentioned, it appears that Dempwolff had doubts about the status of *n' as a PMP phoneme. But it occurred in environments such that it was not phonologically reducible to a palatal allophone of *n (e.g. *n'amuk 'mosquito', *pən'u 'turtle', *n'an'i 'sing'). There is, however, an explanation which would show how *n' was innovated in PMP. I think that a consideration of the phenomenon of nasal grade will provide a good explanation for the origin of *n'.

AN languages normally do not permit consonant clusters. However, most of them do have nasal/stop clusters. These clusters are called nasal grade. There are two types of nasal grade. One is produced by prefixing maŋ- or ŋ- to a stem; this is called nasal substitution and does not occur in Oceanic. The other type of nasal grade is called prenasalisation or simply nasal/stop clusters. In Oceanic prenasalisation occurs both initially and medially, elsewhere it occurs only medially. Neither type of nasal grade occurs in Formosa. The two types of nasal grade have quite different effects.

We will give an example of nasal substitution from Ivatan to show its effects.

maŋ + ili 'village' → maŋili
 maŋ + punu 'full' → mamunu
 maŋ + bətbət 'pluck hairs' → mambətbət
 maŋ + takaw 'steal' → manakaw
 maŋ + kali 'dig' → maŋali
 maŋ + sawud 'knit' → manawud
 maŋ + mata 'raw meat' → maŋmata

These effects occur wherever nasal substitution occurs. A combination maŋ and a vowel has no special effect; maŋ and a voiceless stop combine to give a nasal homorganic to the stop; maŋ and a voiced stop give a homorganic nasal stop cluster; maŋ and s (from *t') give n' in those languages which have palatals and n in those which do not; otherwise maŋ has no special effects. maŋ is an AF affix in the focus system.

Prenasalisation is quite different in its effects and does not appear to have any meaning attached to it. Prenasalisation, except in one case, always produces homorganic nasal/stop clusters. However, this one exception is of great importance historically. In Javanese and Malay the reflex of *n't' is ŋs (Jv laŋse, Ml laŋsai from *lan't'ay 'shade'). As Dahl argues this ŋs must represent an original ŋt' which must be the prenasalised form of *t'. As far as we

know there is no reason to assume that *t' was different from any other stop in relation to prenasalisation so one must assume that all the homorganic nasal/stop clusters were originally ŋ/stop clusters.

There is another curious fact about prenasalisation. In his reconstructions involving the nasal/stop clusters Dempwolff almost always has the nasal in brackets. The brackets indicate that the reflexes of the nasal are not predictable. A language may have a nasal/stop cluster or simply the stop. If we examine the cognate sets we find that one language may have a nasal/stop cluster, another a stop, and a third may have both. It is completely unpredictable. Take for example *ta(ŋ)kub 'to cover' and *da(n)daŋ 'heat'.

Tg	TB	Jv	Ml
takob	takkup	takup	taŋkub
daraŋ/dandaŋ	dadaŋ	dandaŋ	dandaŋ

Such examples are numerous. It is obvious that the presence of the nasal was optional. This means that at one time ŋ could be optionally infixes before the medial stop in a word. However Gonda (1943) attacks the view that you can call this ŋ an affix. Gonda argues that you cannot call something an infix if there is no meaning reconstructible for it. It is true that there is no meaning reconstructible for ŋ. He shows that in many languages, both AN and non-AN, nasals and laterals are infixes to words to make them more expressive and emphatic. In AN *-aɪ- and *-aɣ- occur as infixes with no reconstructible meaning. Therefore Gonda argues that we should reconstruct *-ŋ- as a strek form (a phoneme or group of phonemes added to words to make them more intensive). However, Latta (1978) argues against this stating that merely because one cannot reconstruct the semantics of an affix is not an adequate reason for denying that it was an affix. Therefore it appears that the status of *-ŋ- is somewhat uncertain.

Though nasal substitution and prenasalisation have different effects they appear to be related. From the variation maŋ-~ŋ-, we can suggest that maŋ- could be broken up into ma and ŋ. This ŋ would be the same one found in prenasalisation. The affix maɣ- which is structurally similar to maŋ- can also be broken up in this way as there is evidence for a *-ɣ- infix. This strengthens the idea of splitting maŋ- into ma and ŋ. This is in accord with Dahl (1976:128).

Once we have accepted this relationship, a number of questions arise for consideration. The first is for what level of AN do we wish to reconstruct these processes. This naturally entails determining which languages the processes can be reconstructed for. Secondly, we must explain the difference in effects between the two. Thirdly, we must decide whether prenasalisation occurred originally both initially and medially as in Oceanic or medially only as elsewhere.

The answer to the first question is that we would wish both types of nasal grade for PMP. There are only two forms which provide evidence for prenasalisation in Taiwan. One is the word for star, PAN *bi(t)uqən, Pz bintun, Ss bintö'än, Bun bintoqan 'star' and *qaŋt'u, Bun maqantsuh 'smell of rotten tubers' (Zorc). As homorganic nasal/stop clusters are probably the commonest type of consonant clusters cross-linguistically, we would require a reasonable number of forms before we could exclude the possibility of these forms being due to other factors than prenasalisation. Two forms is certainly not nearly enough to do this. Furthermore, there is no evidence for *maŋ- in Taiwan. If we accept that nasal substitution and prenasalisation are related then we must accept that

nasal substitution is older. Nasal substitution causes much more complex assimilations than prenasalisation which argues strongly that it is older. Secondly, when η was added to ma to form a focus affix it presumably had a much more specific meaning than when it could be infixes in any word base which again means nasal substitution must be older.

Nasal substitution does not occur in Oceanic, but there are traces of it (Reid, personal communication). So one can regard the loss of nasal substitution in Oceanic as being part of the general loss of the focus system. The Taiwanese languages, on the other hand, show no sign of having lost the focus system and no good reason can be proposed for the loss of $ma\eta\text{-}\eta$. Therefore we will regard $ma\eta\text{-}\eta$ as a PMP innovation.

We will assign prenasalisation to PMP. A number of northern Philippine languages such as Ivatan and Bontok do not show any evidence of prenasalisation, neither does Kelabit in Borneo (Blust, personal communication). However, as these languages on other evidence all belong to the PMP subgroup we will, for the moment, regard the absence of prenasalisation as being due to loss. It is possible that the absence of prenasalisation in a number of northern Philippine languages could indicate that prenasalisation was not innovated until later. As we have said, all the evidence indicates that nasal substitution is an older process than prenasalisation and this is the reason for their different effects.

I would say that on the evidence available that it is best to reconstruct prenasalisation medially only. The prenasalised stops which occur initially in Oceanic are probably better regarded as the trace of nasal substitution with η .

Once we accept that nasal substitution is a process innovated at PMP level then we will obtain a perfectly plausible origin for $*n'$. This phoneme would be produced by $ma\eta$ - prefixation to any stem starting with $*t'$. This would produce $*n'$ in all sorts of contrastive environments and would of course be completely undetermined phonologically. This would provide PMP with a $*n'$ phoneme which could then occur in new words added to the lexicon and spread into some old ones.

Therefore, from the phonological data we have the following phenomena which provide conclusive evidence for a PMP subgroup

- a) the split merger of $*n$, $*t$ and $*l$
- b) the creation of $*n'$ by nasal substitution.

We can also add the innovation of prenasalisation to these two pieces of evidence. However, as we indicated earlier, there are some doubts about what subgroup level this should be assigned to so it is not as strong a piece of evidence as the other two.

3.1.2. Grammatical evidence

The grammatical evidence for a PMP subgroup is not as extensive as the phonological evidence. One would have to reconstruct a grammatical system almost the same as the one we have reconstructed simply on the basis of the MP languages. The only language from Formosa which adds something to our reconstruction is Tsou and the Tsou evidence is basically confirmatory.

This is not say, however, that there is no evidence of grammatical innovations in a PMP subgroup. Considering the pronouns first, we find three excellent pieces of evidence for a PMP subgroup. The first of these is the

innovation of *t'i-da as a third plural. In Taiwan *t'i-da occurs only in Amis and there it occurs as a singular. The other two pieces of evidence are provided by Blust (1977). One is the use of *mu as a 2sg gen which does not occur in Taiwan where it is always a 2nd pl. As Blust points out, this is presumably due to politeness and deference restrictions. The third piece of evidence is provided by the oblique pronouns. Outside Taiwan the lpl excl *itən is reflected as *atən apparently under the influence of the lpl excl *amən. These three pieces of evidence provide more conclusive evidence for a PMP subgroup.

Clause structure provides us with one piece of evidence for a PMP subgroup. This is the innovation of *maŋ-ŋ- in the focus system. Blust (1974) provides us with evidence from the construction markers for a PMP subgroup. This is the ligature *ŋ(a). We have reconstructed *a as the PAN ligature and therefore *ŋ(a) must be an innovation.

The final area where one would like to have some evidence for a PMP subgroup is from the lexicon. However, as no studies have been done in this area, we have no evidence. With at least some words we know that they must be PMP lexical additions. These are words with *n' such as *n'amuk 'mosquito', *n'an'i 'sing' and *pən'u 'sea turtle' all of which are fairly basic and occur throughout the MP languages. I suspect that further studies would show much vocabulary which occurs exclusively in the MP languages.

From the evidence presented here, I think that one can conclude that a PMP subgroup of PAN did exist.

3.2. The grouping of the Formosan languages

Now that we have established that MP is a subgroup of AN, we naturally want to know how the Formosan languages group among themselves. The following subgroups have been proposed for the Formosan languages:

- a) Atayalic consisting of Atayal and Sediq;
- b) Tsouic consisting of Tsou, Kanakanabu and Saaroa;
- c) Favorlang-Pazeh-Saisyat;
- d) Amis-Bunun-Siraya-Kavalan;

with the rest being ungrouped.

The Atayalic group seems reasonably well supported by the evidence. Sediq and Atayal appear to have separated fairly recently. Sediq and Atayal have a form mu lsg gen which must be a common innovation since it is not found elsewhere in AN. They also share locative forms of 1st and 2nd sg kənan (s) knan (At) 1st sg, sunan (s+At) 2nd sg. Asai (1953) does not give full paradigms for any other pronouns and so we do not know if Sediq has locative forms for the other pronouns. However, on the basis of the first and second sg forms in Sediq it is highly probable that it does. As we have already mentioned, these locatives are an Atayalic innovation. They both have the PF marker as un indicating a common development of *ə to u. This is all that can be gleaned from the available sources but it appears to be good enough to say that Atayal and Sediq form a subgroup. I suspect that with more information there would be more evidence. Asai's Sediq grammar is rather fragmentary as an overall description of a language.

The evidence for a Tsouic subgroup is much less convincing than that for Atayalic. The phonological innovations which the three Tsouic languages share

can be summarised as:

- a) a merger of *ts and *d and *d' to ts in Kn and ts/s in Saar and Ts;
- b) development of final *an to a (all VC → VCV).

The merger of *ts and *d to *ts is a good argument for subgrouping the three languages because it is such an unusual merger. The development of *an to a is also a reasonable subgrouping argument because all other word-final groups VC acquire a supporting epenthetic vowel in Tsouic.

Grammatically the languages all show a tendency to merge PF and LF. In Tsou -a is the PF marker and in Saaroa it is the optional PF marker. In Saaroa the LF marker is -a(na) and in Kanakanabu it is -a in perf LF. In Kanakanabu imperfective LF and PF are both marked by -ənə. In Kanakanabu perfective PF is marked by -ə. Tsou and Saaroa have both replaced *-ən with -a (as we pointed out earlier this -a can come from PF *-a or LF *-an). On the other hand, Kanakanabu appears to have extended the use of *ən, at least in the imperfective. Therefore we have Tsou-Saaroa grouping against Kn. Kn and Saar merge *d with *d, *d' and *ts whereas Tsou has an r reflex for *d.

Most scholars split Tsouic into Tsou and Kanakanabu-Saaroa. However, there does not appear to be any evidence for this grouping. Tsuchida does not cite any grammatical innovations for a Kanakanabu-Saaroa subgroup. There is a lot of shared vocabulary between Kanakanabu and Saaroa and indeed among all three languages but this does not provide that they form a subgroup. Ferrell (1969:36-39) makes it clear that the Tsouic languages form a cultural group. This would undoubtedly lead to a great deal of borrowing between the languages. It appears therefore that there are considerable doubts about the existence of a Tsouic group.

Tsuchida (1976) suggests that Rukai groups with Tsouic. The evidence he offers is that the Maga and Mantaurean dialects of Rukai have the same final *an to a development as in Tsouic. Like Tsouic, Rukai adds epenthetic vowels to consonant-final stems. These are the only two common innovations. However, the *an to a change does not occur in all Rukai dialects so it cannot be used as a subgrouping argument. The occurrence of the change *an to a in certain Rukai dialects (the ones adjacent to Tsouic) must lessen its value as a subgrouping argument for Tsouic and thereby increase the doubts about the Tsouic subgroup. The innovation of the epenthetic vowel is not a particularly convincing subgrouping argument because it occurs elsewhere in AN as in Sangirese and Palauan. Furthermore it is a fairly natural result of pronouncing final consonants with release.

There is one piece of evidence that Rukai may have subgrouped in Paiwan. In an archaic text concerning headhunting collected by Li, the 2nd pl Category 1 pronoun is mon. This obviously derives from *mun, the PAN Category 3 pronoun. Now Paiwan is the only other AN language which has Category 1 pronouns from PAN Category 3 pronouns. However this one pronoun form is not enough to group Rukai and Paiwan. However if Rukai ever grouped with anything then all traces of this have since been lost by its own innovations as can be seen from the pronoun paradigm (see pronoun paradigms, section 2) and the verb system (see Li 1973:156).

The next subgroup which has been proposed for Formosa is Favorlang, Pazeh, Saisyat (Marsh 1977). The evidence for the subgroup is that all three languages merge *dz (which we do not accept) and *d. In Pazeh and Saisyat these two phonemes merge with *d as well. Unfortunately, there is no other evidence for a subgroup that I have been able to find. The Pazeh and Saisyat pronoun paradigms (set out in section 2) show no evidence of a period of common development. There is no

paradigm available for Favorlang. Nor does it appear that they share any innovations in syntax or CMs. The Pazeh and Saisyat CMs are (Li 1978):

	Pazeh	Saisyat
Topic	ka	∅
Nominative	ki	∅
Accusative	a, u	ka
Genitive	ni	noka, na, ni (pers)
Locative	di	ray, kah
Instrumental	∅	noka
Comitative	∅	ni

I have not altered Li's terminology because the functions of the CMs are not relevant and because I do not have enough information to determine their functions. What is relevant is that the forms of the CMs appear to be unrelated except for the *ni* genitive which is PAN. Again I have not been able to obtain any Favorlang forms. From Marsh's description, Pazeh, Saisyat and Favorlang verbal systems do not appear to show any common innovations. The Favorlang focus affixes are *m* AF, *-en* PF, *-anvayan* LF, *i-* IF, *-in-* perf; in Pazeh these are *mv-* (v homorganic to first vowel in stem) AF, *-un* PF, *-an* LF, *si:-* IF and *-in-* perf; and in Saisyat these are *-om-*, *m-* AF, *-in* PF, *-an anai* LF, *si-* IF and *-in-* past. They all have *pa* causative. In all the languages perfective PF constructions have only *-in-* and no **-ən*. These things are all reconstructible for PAN and naturally therefore cannot be subgrouping arguments. From the information available at the moment, there is only one innovation, a merger, to support the existence of a Pazeh-Saisyat-Favorlang subgroup. Therefore we must regard the status of this subgroup like that of Tsouic as questionable.

Finally we will consider the Amis-Bunun-Kavalan-Siraya (Ferrell 1979). These languages all merge **t* and **ts* to *t*. Amis and Kavalan have merged medial **g'* and **n* to *n*. Bunun and Kavalan have merged **n* and **ɲ* to *n*. Amis has merged **ɟ* with **t* and **ts*. Amis has an accusative CM *tu ... (an)* (Reid 1971). The non-personal accusative in Kavalan is *tu* (Li 1978). These two CMs must be cognate. This CM also occurs in Thao (Li 1978). Thao merges **t* and **ɟ* to *t* just like Amis. Thao shares the irregular genitives *nak* and *nam* with Bunun. Thao and Kavalan have **a* as a ligature with the following variants, *a* after consonants, *wa* after back vowels and *ya* after front vowels. This suggests some sort of common development since PAN. Now these facts all point towards Thao having some sort of relationship with the subgroup we have proposed though it lacks the **t* and **ts* merger which defines the group. However, if we examine the mergers which we have been considering we find the following facts.

- *n* and **ɲ* merged in *n* in Kv, Bun and Kn.
- *t* and **ts* merged in Am, Bu, Kv, Sir.
- *d* and **ɟ* merged in Bun, Th, Kn, Pz, Ss, At, Se, Sar.
- *t* and **ɟ* merged in Am, Th.
- *d* and **ts* merged in Th, Ts, Kn, Sar.

All these mergers involve continuous areas except for the Amis and Thao merger of **t* and **ɟ* to *t*. The impression one gets when examining these mergers is that a wave model would be a far more accurate model of the situation than a tree diagram. The only mergers which are really convincing as subgrouping evidence are Thao and Amis **t* and **ɟ* to *t* and the merger of **ts* and **d* to *ts* in the Tsouic languages. These mergers are unusual and therefore argue for some common period of development. The Favorlang-Pazeh-Saisyat group looks weaker than ever if we consider the fact that Pazeh-Saisyat and Bunun all merge **d*,

and *ɖ to d (or r in Saisyat). Amis and Thao share the unusual *t and *ɖ merger to t, which cannot be areal, and the accusative marker tu. However this last thing occurs elsewhere. It seems that we will be able to subgroup languages only on the basis of a large number of shared similarities. At the moment it seems that there is a case for subgrouping Atayal and Sediq into Atayalic and much less convincing arguments for subgrouping Saaroa, Kanakanabu and Tsou into Tsouic and Amis and Thao.

The problems we have in subgrouping the Formosan languages can have two explanations. One is that they all form one subgroup, or at least the non-Atayalic-Tsouic languages form a subgroup. The only piece of evidence I know of for a Formosan subgroup is the fact that *d and *d' are always reflected the same way in the languages of Formosa. If one examines this merger more closely, one finds that this does not provide any evidence for a subgroup. In Paiwan *d and *d' merge to d' (i.e. d becomes d'). In Siraya *d and *d' merge to d with the merger going in completely the opposite direction. In Amis *d, *d', *l, *ɣ all merge to an alveolar tap so we have no idea whether *ɖ and *d' ever were merged by themselves. In Kanakanabu *ts, *d, *d', *dz and *ɖ all merge to ts so again we have no idea whether *d and *d' ever had an independent merger. The same can be said for Pazeh and Saisyat where *d', *d, and *ɖ all become d and r respectively. Apart from these problems this merger would not uniquely identify the Formosan languages anyway. The languages of the Philippines all merge *d and *d' as well. Therefore we would have to subgroup the Philippines and Formosa which goes against the much better evidence we have presented for PMP. There does not seem to be convincing or even vaguely suggestive evidence that there is a Formosan subgroup. We will consider an alternative solution in the next section.

3.3. The relationship between the PMP subgroup and the Formosan languages

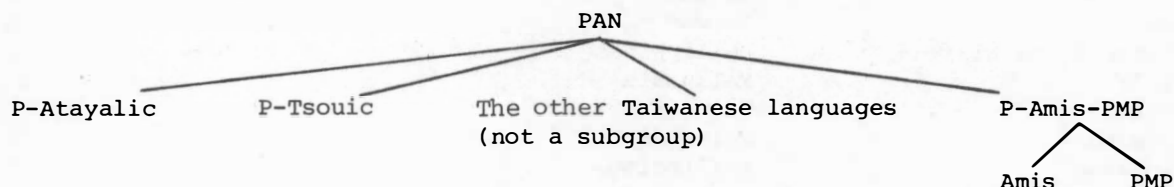
The question now is whether PMP is a first-order subgroup of PAN, whether it groups with one of the Formosan subgroups we have suggested or one of the ungrouped Formosan languages. PMP does not group with Atayalic or Tsouic. It shares no innovations with these subgroups. Nor does it appear to share any innovations in common with most of the ungrouped languages.

However, it does share the mergers of *t and *ts to t with Amis, Bunun, Siraya and Kuvulan. Of these languages, PMP appears to have the most similarities with Amis. Amis has the ka pronouns which the other three languages do not. The ka pronouns are the dominant forms in PMP; the non-ka pronouns occur only rarely in some Philippine languages. Amis is the only Taiwanese language with a reflex *t'i-da which we have shown to be an innovation. This common innovation of *t'i-da is an excellent argument for subgrouping PMP and Amis. However, there does not appear to be any argument for assigning any other language to this Amis PMP subgroup.

Therefore we have Amis PMP and Atayalic with a doubtful Tsouic subgroup and the rest of the Taiwanese languages being subgrouped. This interpretation of Austronesian subgrouping can have only one meaning historically. This is that Taiwan is the homeland of the Austronesians. High internal diversity among the Taiwanese languages combined with the fact that all the non-Taiwanese languages form a subgroup which can in turn be grouped with Amis, compel this conclusion (Sapir 1949b, Bloomfield 1933). This conclusion is in line with the archaeological and anthropological evidence (Bellwood 1978).

This would explain why it is so difficult to subgroup the Taiwanese languages. If Taiwan is the homeland then the Taiwanese languages have been interacting for at least 5,000 years (the Malayo-Polynesian departure from Taiwan must go back 5,000 years). Greater than 5,000 years of interaction between related languages is enough time to obliterate any original subgroup in Taiwan. As we have said, if we have Taiwan as the homeland with one group of people the Malayo-Polynesians leaving Taiwan to settle the rest of Austronesia then we would expect the Taiwanese languages to be as different from one another as they are from PMP.

Therefore we propose the following PAN family tree:



NOTE

List of languages mentioned, with abbreviations as used in the text.

Language	Location	Abbreviation
Acehnese	Sumatra	
Amblau	Indonesia	
Amis	Taiwan	Am
Atayal	Taiwan	At
Bikol	Philippines	Bk
Bisayan	Philippines	
Buginese	Sulawesi	
Bunun	Taiwan	Bun
Cebuano	Philippines	
Fijian	Fiji	Fj
Gayo	Sumatra	
Hiligaynon Bisayan	Philippines	Hl Bs
Ibanag	Philippines	
Ivatan	Islands between Philippines and Formosa	Iv
Javanese	Java	Jv
Kanakanabu	Taiwan	Kn
Kavalan	Taiwan	Kv
Kelabit	Borneo	Kel
Madurese	Madura	Md
Magori	Papua	
Malagasy	Madagascar	Mal
Malay	Malaya	Ml

Language	Location	Abbreviation
Ngadju Dayak	Borneo	NgD
Nias	Islands off Sumatra	
Okolod Murut	Sabah	
Paiwan	Taiwan	Pw
Palauan	Palau Islands	Pl
Pazeh	Taiwan	Pz
Puyuma	Taiwan	Pu
Rukai	Taiwan	Ruk
Saaroa	Taiwan	Sar
Saisyat	Taiwan	Ss
Samar Leyte Bisayan	Philippines	SL Bs
Sa'a	Melanesia	Sa
Sediq	Taiwan	Se
Siraya	Taiwan	Sir
Subanon	Philippines	
Sundanese	Western Java	
Tagalog	Southern Luzon (Philippines)	Tg
Thao	Taiwan	Th
Toba Batak	Sumatra	TB
Tongan	Tonga	To
Tsou	Taiwan	Ts
Western Bukidnon Manobo	Philippines	WBM

Language Groups

Malayo-Polynesian - the non-Formosan Austronesian languages	MP
Oceanic - the languages of Micronesia Melanesia and Polynesia	OC
Philippine - the languages of the Philippines	Ph

Proto-Languages

Proto-Amblau	Indonesian	
Proto-Austronesian		PAN
Proto-Malayo-Polynesian		PMP
Proto-Oceanic		POC
Proto-Philippine		PPh

BIBLIOGRAPHY

ANDERSEN, Henning

- 1973 Abductive and deductive change. *Language* 49:765-793.

ANTTILA, Raimo

- 1972 *An introduction to historical and comparative linguistics*. New York: Macmillan.

ARMS, David G.

- 1974 Transitivity in Standard Fijian. Ph.D. dissertation, University of Michigan.

ASAI, Erin

- 1936 *A study of the Yami language, an Indonesian language spoken on Botel Tobago island*. Leiden: Burgersidijk & Niermans.
1953 *The Sedik language of Formosa*. Kanazawa, Japan: Cercle Linguistique de Kanazawa, Kanazawa University.

BELLWOOD, Peter

- 1978 *Man's conquest of the Pacific*. New York: Oxford University Press.

BLOOMFIELD, Leonard

- 1933 *Language*. New York: Henry Holt; London: Allen & Unwin.

BLUST, Robert A.

- 1974 Proto-Austronesian syntax: the first step. *Oceanic Linguistics* 13:1-15.
1976 Review of Otto Chr. Dahl *Proto-Austronesian*. *Language* 52:221-237.
1977 The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics, University of Hawaii (WPLUH)* 9/2:1-15.
1978 *The Proto-Oceanic palatals*. Memoir 43. Auckland: The Polynesian Society.

CHEN, Matthew Y. and William S.-Y. WANG

- 1975 Sound change: actuation and implementation. *Language* 51/2:255-281.

CHEN, Teresa M.

- 1982 Verbal construction and verbal classification in Nataoran-Amis. Ph.D. dissertation, University of Hawaii.

DAHL, Otto Chr.

- 1976 *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph 15. Lund: Studentlitteratur.

DARDJOWIDJOJO, Soenjono

- 1978 Nominal derivation in Indonesian. In: Wurm and Carrington, eds 1978: 503-528.

DELBRÜCK, Berthold

- 1884 *Einleitung in das Sprachstudium*. Zweite Auflage. Leipzig: Breitkopf und Härtel.

DEMPWOLFF, Otto

- 1934- *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3 vols.
1938 Berlin: Reimer.

DIXON, R.M.W.

- 1979 Ergativity. *Language* 55:59-138.

DYEN, Isidore

- 1951 Proto Malayo-Polynesian *Z. *Language* 27:534-540.
1956 The Ngaju-Dayak 'old speech stratum'. *Language* 32:83-87.
1974 The Proto-Austronesian enclitic genitive pronouns. *Oceanic Linguistics* 13:17-31.

EGEROD, Søren

- 1966 Word order and word classes in Atayal. *Language* 42:346-369.

EMENEAU, M.B.

- 1956 India as a linguistic area. *Language* 32:3-16.

FERRELL, Raleigh

- 1969 *Taiwan Aboriginal groups: problems in cultural and linguistic classification*. Academia Sinica Monograph 17. Taipei: Institute of Ethnology.
1979 Phonological subgrouping of Formosan languages. In: Naylor, ed. 1979:241-254.
1982 *Paiwan dictionary*. PL, C-73.

FOLEY, William A.

- 1976 Comparative syntax in Austronesian. Ph.D. dissertation, University of California, Berkeley.

GAUCHAT, Louis

- 1905 l'unité phonétique dans le patois d'une commune. *Aus romanischen Sprachen und Literaturen: Festschrift Heinrich Morf*, 175-232. Halle: Max Niemeyer.

GIVÓN, Talmy

- 1975 Topic, pronoun, and grammatical agreement. In: C.N. Li, ed. *Subject and topic*, 149-188. New York: Academic Press.

GONDA, J.

- 1943 Inwendige nasaal- en liquida-verbindingen in Indonesische talen. *Bijdragen tot de taal-, land- en Volkenkunde (BijdrTLV)* 101:141-206.

HARMON, Carol M.

- 1974 Reconstructions of Proto-Manobo pronouns and case marking particles. *WPLUH* 6/6:13-46.

HAUDRICOURT, André G.

- 1965 Problems of Austronesian comparative philology. *Lingua* 14:315-329.

HENDERSON, Eugénie J.A.

- 1965 The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15:400-434.

JENG, Heng-hsiung

- 1977 *Topic and focus in Bunun*. Academia Sinica Special Publication 72. Taipei: Institute of History and Philology.

JOSEPHS, Lewis

- 1975 *Palauan reference grammar*. Honolulu: The University Press of Hawaii.

KEENAN, Edward L.

- 1976 Towards a universal definition of subject. In: C.N. Li, ed. *Subject and topic*, 303-333. New York: Academic Press.

KRISHNAMURTI, Bh.

- 1978 Areal and lexical diffusion of sound change: evidence from Dravidian. *Language* 54/1:1-20.

LABOV, William

- 1963 The social motivation of a sound change. *Word* 19:273-309.
- 1966 *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- 1972 The social stratification of r in New York City department stores. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- 1974 On the use of the present to explain the past. In: L. Heilmann, ed. *Proceedings of the Eleventh International Congress of Linguists...* Bologna, 1972, vol.2:825-851.

LATTA, Christian

- 1978 Internal reconstruction. Ph.D. dissertation, Ohio State University.

LI, Paul Jen-kuei

- 1973 *Rukai structure*. Academia Sinica Special Publication 64. Taipei: Institute of History and Philology.
- 1975 *Rukai texts*. Academia Sinica Special Publication 64/2. Taipei: Institute of History and Philology.
- 1978 The case marking systems of the four less-known Formosan languages. In: Wurm and Carrington, eds 1978:569-615.

MARSH, Mikell A.

- 1977 Favorlang-Pazeh-Saisiat: a putative Formosan subgroup. Ph.D. dissertation, Washington State University, Pullman.

MARTINET, André

- 1952 Function, structure, and sound change. *Word* 8:1-32.

NAYLOR, Paz Buenaventura

- 1975 Topic, focus, and emphasis in the Tagalog verbal clause. *Oceanic Linguistics* 14/1:12-79.

NAYLOR, Paz Buenaventura, ed.

- 1979 *Papers from the Second Eastern Conference on Austronesian languages*. Ann Arbor: University of Michigan.

PAWLEY, Andrew and Lawrence A. REID

- 1976 The evolution of transitive constructions in Austronesian. *WPLUH* 8/2:51-74.

REID, Lawrence A.

- 1971 *Philippine minor languages: wordlists and phonologies*. *Oceanic Linguistics* Special Publication 8. Honolulu: University of Hawaii Press.
- 1974 The pronominal systems of Proto-Cordilleran, Philippines. *WPLUH* 6/6:1-12.
- 1978 Problems in the reconstruction of Proto-Philippine construction markers. In: Wurm and Carrington, eds 1978:33-66.
- 1979 PAN genitive alternation: the Philippine evidence. *WPLUH* 11/2:45-55.

SAPIR, Edward

- 1949a The psychological reality of phonemes. In: D.G. Mandelbaum, ed. *Selected writings of Edward Sapir in language, culture and personality*, 46-60. Berkeley: University of California Press.
- 1949b Internal linguistic evidence suggestive of the northern origin of the Navaho. In: Mandelbaum, ed. 1949:213-224.

SCHACHTER, Paul

- 1977 Reference-related and role-related properties of subjects. In: Peter Cole and Jerrold M. Sadock, eds *Syntax and semantics*, vol.8: *Grammatical relations*, 279-306. New York: Academic Press.

SCHACHTER, Paul and Fe T. OTANES

- 1972 *Tagalog reference grammar*. Berkeley: University of California Press.

SCHUCHARDT, Hugo

- 1885 *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: Robert Oppenheim. Reprinted and translated in Vennemann and Wilbur, eds 1972:1-37, 39-72.

SILVERSTEIN, Michael

- 1976 Hierarchy of features and ergativity. In: R.M.W. Dixon, ed. *Grammatical categories in Australian languages*, 112-171. New Jersey: Humanities Press; Canberra: Australian Institute of Aboriginal Studies.

SMITH, Neilson V.

- 1973 *The acquisition of phonology: a case study*. London: Cambridge University Press.

STURTEVANT, Edward H.

- 1917 *Linguistic change*. Chicago: University of Chicago Press.

TRUDGILL, Peter

- 1971 *The social differentiation of English in Norwich*. Ph.D. dissertation, University of Edinburgh.

TSUCHIDA, Shigeru

- 1976 *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph 5. Tokyo: Toyo Shuppan.

TWADDELL, William F.

- 1935 *On defining the phoneme*. Language Monograph 16. Baltimore.

VENNEMANN, Theo

- 1972 *Phonetic analogy and conceptual analogy*. In: Vennemann and Wilbur, eds 1972:181-204.

VENNEMANN, Theo and Terence H. WILBUR, eds

- 1972 *Schuchardt, the Neogrammarians, and the transformational theory of phonological change*. Linguistische Forschungen 26. Frankfurt: Athenäum.

WANG, William S.-Y.

- 1969 *Competing changes as a cause of residue*. Language 45/1:9-25.

WOLFF, John U.

- 1973 *Verbal inflection in Proto-Austronesian*. In: Andrew Gonzalez, ed. *Parangal kay Cecelio Lopez*, 71-91. Philippine Journal of Linguistics Special Monograph 4. Quezon City.

- 1974 *Proto-Austronesian *r and *d*. Oceanic Linguistics 13:77-121.

WURM, S.A. and Lois CARRINGTON, eds

- 1978 *Second International Conference on Austronesian Linguistics: proceedings*. PL, C-61. (2 fascicles). [SICAL].

ZORC, R. David

- 1977 *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. PL, C-44.

