

In Defense of Nuclear Austronesian (and Against Tsouic)*

Malcolm Ross

The Australian National University

Ross (2009) proposed the Nuclear Austronesian hypothesis, whereby Puyuma, Tsou and Rukai are each single-member first-order subgroups of Austronesian and all other Austronesian languages belong to a Nuclear Austronesian subgroup. The basis of this subgrouping is a complex innovation whereby certain Proto Austronesian nominalizers came also to mark indicative verbs. This paper falls into two parts. The first surveys kinds of evidence that historical linguists use in subgrouping and proposes metrics (§2) that are then applied to the innovations that support Nuclear Austronesian (§3) and other recent first-order subgroupings of Austronesian (principally Formosan) languages (§4). The second part argues that the commonly accepted Tsouic subgroup, which is incompatible with the Nuclear Austronesian hypothesis, is not supported by the evidence. Instead it reflects long-term contact between Tsou on one hand and Kananavu and Saaroa on the other (§5). In conclusion, it is tentatively suggested that the southern part of the Taiwan highlands appears to be the oldest Austronesian homeland area.

Key words: Austronesian, Nuclear Austronesian, Formosan, Tsouic, Tsou, subgrouping

1. Introduction

In Ross (2009) I proposed the Nuclear Austronesian (NAn) hypothesis, whereby Puyuma, Tsou and Rukai are each single-member first-order subgroups of Austronesian

* I am very much indebted to Stacy Fang-ching Teng, who read and summarized Chinese-language articles for me. I am also grateful to Robert Blust and Peter Trudgill for answering my questions during the preparation of an earlier draft of parts of this paper. The comments of two anonymous referees afforded considerable assistance in my final revision, even though neither of them will agree with all my conclusions. Much of the research on which this paper is based was conducted during periods of research at the Institute of Linguistics of Academia Sinica (ILAS), Taipei, from September to December 2003 and from September 2008 to January 2009. During the earlier period I was a Visiting Research Fellow under the project ‘A morpho-syntactic study of Rukai’ (Principal Investigator: Elizabeth Zeitoun), supported by National Science Council of Taiwan grant NSC92-2811-H-001-010. I am grateful to Elizabeth Zeitoun, Ho Dah-an and Jackson Sun for facilitating these visits and to the NSC and ILAS for their support.

and all other Austronesian languages belong to a NAn subgroup. My purpose in this paper is to defend the NAn hypothesis, which is outlined below in §3.

The defense has two parts. First, I endeavor to show why three other high-order subgroupings of Austronesian that have been proposed in the last fifteen years, Ho (1998), Blust (1999) and Sagart (2004),¹ do not offer accounts of the early history of Austronesian that are as convincing as the NAn hypothesis. Second, I argue that there is no Tsouic subgroup. Tsouic, consisting of Tsou, Kanakanavu and Saaroa, has formed a part of all previous high-order Austronesian subgroupings. However, Tsouic is incompatible with the NAn hypothesis, under which Tsou is a single-member first-order subgroup whilst Kanakanavu and Saaroa are members of the NAn subgroup.²

The four subgroupings mentioned above agree that all Austronesian languages outside Taiwan belong to a single subgroup, Malayo-Polynesian, and that the Formosan languages (the Austronesian languages of Taiwan) belong to more than one primary subgroup of Austronesian. But just how the Formosan languages should be subgrouped is a matter on which the four subgroupings disagree, and it is important to consider how these disagreements have arisen before addressing the disagreements themselves, since the facts on which they are based require explanation (or must be shown not to be facts). I address these matters in §2.

2. Matters of method

A linguist applying the comparative method identifies innovations that are shared by the member languages of a subgroup because they had occurred in these members' shared ancestor. Disagreements about subgrouping are either about the evidentiary value of various kinds of shared innovation or about data interpretation, i.e. whether a particular innovation actually occurred.

Part of the evidentiary value of an innovation centres on the question of whether the presence of an innovation in a language is the outcome of its **transmission** from one

¹ Earlier subgroupings are surveyed by Blust (1999).

² Abbreviations of names of protolanguages are PAN (Proto Austronesian), PAta (Proto Atayalic), PMP (Proto Malayo-Polynesian), PNAn (Proto Nuclear Austronesian), PRuk (Proto Rukai). NAn is used for Nuclear Austronesian. In Table 5 and in cognate sets names of Formosan languages are abbreviated as follows: **Amis**, Mayrinax **Atayal**, **Babusa**, **Basay**, **Bunun**, **Favorlang**, **Hoanya**, **Kanakanavu**, **Kavalan**, **Paiwan**, **Papora**, **Pazih**, Nanwang **Puyuma**, **Saaroa**, **Saisiyat**, **Seediq**, **Siraya**, **Taokas**, **Thao**, **Trobiawan**, **Tsou**. Rukai dialect abbreviations are **RukBud** (Budai), **RukMag** (Maga), **RukMan** (Mantauran), **RukTan** (Tanan), **RukTon** (Tona). The symbol † is used to mark constructed forms that are expected but do not in fact occur. SPR = Starosta, Pawley & Reid (see §3.1).

generation to the next or of its **diffusion** across a speech community boundary (Labov 2010:307-309). That is, is the innovation the result of inheritance or of contact? This distinction plays a major role in the argument of this paper.

2.1 Foundational matters in Formosan linguistic prehistory

Two foundational points apply to Formosan linguistic prehistory. Firstly, claims about highest-order subgroupings are unavoidably weaker than claims about lower-order subgroupings. Reconstruction at any lower node in the genealogical tree, e.g. Proto Malayo-Polynesian (PMP—the ancestor of all Austronesian languages outside Taiwan), can call on evidence both from languages below that node (internal evidence) and from languages descended from sister and more distantly related nodes (external evidence). A shared innovation that depends on the reconstruction of a particular feature in the protolanguage entails some circularity, as the output of the innovation is part of the data on which the reconstructed feature is based. I call this the ‘root-node limitation’.

The second foundational point concerns the linguistic geography of Austronesian Taiwan and the role of subgrouping in reconstructing Formosan linguistic prehistory. We can be reasonably confident that Austronesian languages have been spoken in Taiwan since the time of the Tapengkeng Culture around 3500 BC (Bellwood et al. 2011). Until about 2000 years ago, Formosan communities were Neolithic and formed smallscale agricultural societies like those typical of much of the Austronesian speaking world before modernization. Their initial spread across Taiwan can be modelled on the basis of Neolithic cultures elsewhere. After speakers of (pre-)PAn first settled in Taiwan and established rice-growing settlements, their population gradually increased and they eventually occupied all the agriculturally suitable areas of the island, in the process probably absorbing hunter-gatherer communities of the pre-Austronesian Changpin Culture. Their expansion resulted in many small independent communities. Linguistically this was reflected in the diversification of PAn into a network first of dialects and then of languages. But communities often remained in contact, with the consequence that innovations in one language would sometimes spread to its neighbor.³ The resulting pattern is one whereby languages typically do not fall into subgroups, each defined by a set of coterminous shared innovations, but instead form one or more linkages whose

³ About 2000 years ago some groups acquired iron. It is difficult to know what effect this might have had on speech community size, but there were still at least twenty, and perhaps more, Austronesian languages when the Southern Min speaking population started to arrive in the seventeenth century.

member languages display an overlapping pattern of innovations.⁴ The mode of language diversification was thus what Pawley & Green (1984), writing about the Austronesian languages of Oceania, describe as ‘network-breaking’ rather than ‘radiation’ (Pawley 1999:131). Indeed, Pawley suggests that network-breaking was the usual mode of diversification throughout much of the history of Austronesian.

I suggest below that both subgroups and linkages can be recognized in Formosan linguistic prehistory. Both have been identified elsewhere in the history of Austronesian. For example, Malayo-Polynesian (Blust 1977) and Oceanic (Dempwolff 1937) are large subgroups, while Central Pacific (Geraghty 1983) is a linkage. An account that recognizes both will be more faithful to the data than an application of the family tree model which ‘treats contact or “wave model” effects as disturbing elements that limit the precision of the reconstruction’ (Labov 2010:308). This prevents us from drawing a fully articulated family tree, but such a tree would distort the events implied by the data.

2.2 Evidence of subgroups and evidence of linkages

To detect a distinction between subgroups and linkages we need to know how the evidence for each differs. The principal difference is distributional, and types of innovation play a supporting role, as the discussion in §2.2.2 explains.

2.2.1 Distributional evidence

The difference in distribution was foreshadowed above. A subgroup is ideally defined by a set of coterminous⁵ innovations. For example, Malayo-Polynesian, the subgroup containing all Austronesian languages spoken outside Taiwan, is defined by the set of coterminous innovations which include those in (1).

- (1) a. Phonological:
 - i. Merger of PAn *S and *h as PMP *h.
 - ii. PAn *L > PMP *l word-initially and *n word-medially and -finally.
- b. Lexical: PAn *Siwa ‘nine’ > PMP *siwa (for expected *hiwa) (Blust 1995b); PAn *biRbiR ‘lips’ > PMP *bibiR (Blust 1995a:630)
- c. Morphological:
 - i. The politeness shift in PMP pronouns (Blust 1977, Ross 2006).

⁴ This sense of the term ‘linkage’ was introduced into Austronesian historical linguistics in Ross (1988:8 and *passim*).

⁵ That is, having the same boundary.

- ii. Introduction of the PMP actor voice dynamic verb prefixes *maN- and *maR-, the actor voice aptative/potential prefix *maka-, and the actor voice ‘social’ prefix *maki-, together with the corresponding undergoer voice and nonfinite prefixes *paN-, *paR-, *paka- and *paki- (Blust 2003b, Kaufman 2009, Liao 2011a, 2011b).⁶

Blust (1999) sometimes refers to innovations that are only partially complete. For example, he lists truncation of the diphthongs *-ay and *-aw (i.e. deletion of *-y and *-w) among the innovations characterizing the Western Plains group. Truncation is coterminous in Favorlang-Babuza, Papora and Hoanya, but in Thao it is incomplete, and thus sub-coterminous in the subgroup as a whole. This situation has at least two possible explanations. One is that truncation was ongoing when Proto Western Plains broke up, but went to completion in the shared ancestor (Proto Central Western Plains) of the four languages that reflect completion. Another is that truncation occurred only in Proto Central Western Plains, and then spread by contact from a Central Western Plains language to Thao. If the latter explanation is true, then glide truncation is not evidence for the membership of Thao in the Western Plains subgroup—and sub-coterminous innovations are not convincing evidence of a subgroup.

Whereas the innovations that define a subgroup are coterminous, those that connect the languages of a linkage typically form an overlapping pattern where, for example, languages A and B share innovations 1, 2 and 3, languages C and D share innovations 2, 3 and 4, and languages D and E share innovations 2, 4 and 5 etc. The innovations in (2), extracted from Table 5, form such a pattern.

(2)	Pazih	Saisiyat	Atayalic	Amis	Kavalan	PMP
*q > Ø	yes				yes	
*q > ʔ		yes		yes		
*t/*C				yes	yes	yes
*d/*z	yes	yes	yes	yes	yes	
*S/*x ⁷				yes	yes	yes
*s/*x		yes	yes		yes	
*j/*n				yes	yes	

This kind of pattern is usually the result of related languages being in contact, i.e. of there being speakers who are bi- or multilingual in the languages, so that bilingual

⁶ These prefixes represent refunctionalizations and functional extensions of earlier affix sequences.

⁷ PAn *x is Tsuchida’s (1976) *S₂. See §4.1 and §5.2.7 below.

speakers copy features from one language into another.⁸ The newly introduced features are contact-induced innovations. Note that although Amis and Kavalan share more innovations in (2) than any other language pair, this is scarcely evidence of subgrouping, since the innovations fail the ‘no independent innovation’ condition introduced in §2.2.2.

2.2.2 Types of innovation

An innovation has evidentiary value if it is reasonably certain (a) that it has not occurred independently in the different members of the subgroup and (b) that it has not been copied across language boundaries. I refer to these conditions as the ‘no independent innovation’ and ‘no copying’ conditions. If there is a likelihood that it fails one of them, then its strength as a subgroup-defining innovation is reduced. Of course, if it fails the ‘no copying’ condition, then it may still be evidence of a linkage.

In a paper that attempts to put subgrouping on a mathematical foundation, Ringe, Warnow & Taylor (2002) rank morphological innovations highest on the scale of evidentiary value, phonological innovations somewhere in the middle, and lexical replacement innovations lowest. I discuss the three types below in this order, and introduce a somewhat finer categorization.

2.2.2.1 Morphological (and syntactic) changes

Morphological and syntactic innovations are considered together here because they signal different things: they tend to be evidence respectively of transmission and diffusion.

By morphological change I mean change in the form and/or function of bound morphemes. These changes may entail the addition of new morphemes to an existing paradigm, like the innovation of the new PMP verbal prefixes listed in (1c-ii),⁹ or they may be idiosyncratic changes in individual morphemes, like loss of PAN *S- from PAN *Si- ‘circumstance nominalizer, circumstance undergoer voice’ in its PMP reflex *i-, for expected *hi- (Blust 1995a:627). Ringe et al. (2002:68) regard changes in inflectional

⁸ I prefer the term ‘copying’ to ‘borrowing’, because ‘borrowed’ items are often inexact copies, especially where calquing is involved (Ross forthcoming).

⁹ These prefixes did not arise *ex nihilo*. PMP *maR-/paR- reflects the PAN stative reciprocal prefix reconstructed by Zeitoun (2002). Kaufman (2009) suggests that *maN-/paN- may reflect *ma-/pa- ‘verbalizer’ + *ŋ ‘plurality of object or action’ (it retains a distributive function in the Philippine language Binukid; Post 1992). *-N- indicates homorganic nasal substitution (Blust 2004) which occurs infrequently and unpredictably in Formosan languages (Tsuchida 1976:257-258, Starosta 2002:194-195, Wolff 2007). The PMP innovation in each case is the emergence of a prefix with new function(s). For further discussion see Ross (2012).

morphemes as more reliable indicators of subgrouping than sound changes ‘because inflectional systems are such tightly integrated idiosyncratic constructs that conditions which would give rise to similar changes are unlikely to recur in different languages’, i.e. they are likely to fulfil the ‘no independent innovation’ condition. Because of this tight integration, they are also more likely to fulfil the ‘no copying’ condition. Even Thomason & Kaufman (1988:52, 98), one of whose goals is to show that, given appropriate conditions, anything can be copied, agree that inflectional morphemes are less likely to be copied than other kinds of item and follow Weinreich (1953:32) in suggesting that inflectional copying depends on the typological similarity of the donor and recipient languages.

Because bound morphemes are typically formants in words that belong to a paradigm, sometimes changes in bound morphemes are a dimension of change that affects a whole paradigm. Such a change is the nominalization-into-verb innovation of NAn (whereby nominalizations replaced older verb forms in many independent declarative and interrogative clauses, with both morphological and syntactic consequences; see §3.1). Another paradigmatic change is the PMP elaboration of the PAn system of morphological verb classes with the new affixes listed in (1c-ii) and the accompanying expansion of the NAn system of voices and applicative-like affixes (Ross 2012). Changes like these constitute an integrated set of morphological innovations, often with syntactic consequences, so that the improbability of independent parallel innovation or copying noted by Ringe et al. with regard to single morphological changes is multiplied and approaches absolute improbability.

Paradigmatic changes usually occur only during transmission. The effects of diffusion are quite distinct, precisely because the forms of bound morphemes are typically not copied. Instead, diffusion is sometimes signalled by changes in syntax, as bilingual speakers calque constructions from one of their languages into the other. This eventually results in copying of syntax, such that Subject-Verb-Object is flipped to Subject-Object-Verb and prepositions are replaced by postpositions, as has happened in Oceanic languages of New Guinea as a result of contact with Papuan languages (Ross 2007). One might expect that in such circumstances speakers would also copy bound morphemes from language to language, but this happens only rarely. François (2011) shows that even where bilinguals’ two languages are closely related and structurally isomorphic, bound morphemes are not copied.

Morphological changes, then, do not usually signal contact and are very likely to satisfy the ‘no copying’ condition. They are also likely to fulfil the ‘no independent innovation’ condition, except where parallel grammaticization occurs. Where there are paradigmatic changes of the kind found in NAn and Malayo-Polynesian, their very complexity means makes it is virtually certain that they satisfy both conditions.

2.2.2.2 Sound change

Blevins (2004) accounts for alleged phonological universals diachronically. That is, the shared structural properties of phonologies across languages and language families are the outcomes of universal tendencies in articulation and perception which cause different languages to undergo similar sound changes. If she is right—and it appears that she is—then ‘individual changes are usually so “natural” that they can easily be repeated in different lines of descent’ and ‘their probability of parallel development is thus relatively high for most apparently shared sound changes, and the probability of historically shared development is correspondingly low’ (Ringe et al. 2002:66-67). In other words, the probability that sound changes satisfy the ‘no independent innovation’ condition is not high.

There is also evidence that they sometimes fail the ‘no copying’ condition too. A sound change has a distributional parameter that is important for subgrouping, its distribution across its potential target words in the lexicon. A number of sound changes in Formosan languages are partial, i.e. they do not apply to all their potential targets. Does a partial sound change reflect transmission or diffusion?

This question has a controversial history. Labov (2010:260-261, 285-286) talks about two theories of transmitted sound change. In Neogrammarian theory, sound change is ‘a phonetically driven process that affects all words in a phonologically defined set’ (Labov 2010:285). In the second theory, ‘lexical diffusion’, a sound change diffuses through the lexicon on a word-by-word basis (Chen & Wang 1975:257), affecting more frequently used words first (Bybee 2002). Labov finds on the basis of examples from his own research that Neogrammarian theory represents the normal outcome of transmission, whilst ‘lexical diffusion’ results from contact—from adult speakers adopting a sound change from a neighboring dialect or language (Labov 2010:307-312, 347). They copy their neighbors’ pronunciation of frequently used words, thereby acquiring the outcome of a sound change that has occurred in the neighbors’ speech variety without applying the change to all its potential targets. The neighbors’ sound change is acquired with a loss of regularity and generality. Thus a partial sound change reflects diffusion (Blevins 2004:274).

One explanation of the partial application of diphthong truncation in Thao (§2.2.1, §4.1) is that it is a change of this kind—evidence of diffusion from a Western Plains into Thao, not of subgrouping.

Labov’s transmission/diffusion dichotomy is an oversimplification, however. In smallscale Neolithic societies children often grow up bilingually, and sometimes the psychologically dominant language is not the heritage language emblematic of the community’s identity but the language of another group with whom they have frequent

and intense interaction (its speakers may have married into the village, for example). Bilingual speakers may copy features of the psychologically dominant language into the heritage language. If this happens frequently among preadolescent children, the features become part of the heritage language and are passed on to the next generation. These features may include sound changes. Now sound changes diffused via adult speakers are acquired via individual lexical items, not as rules, but preadolescents have no difficulty acquiring rules (Ross forthcoming). Thus if a sound change is acquired before adolescence, it will be as regular in application (and as complex in conditioning) as a change that occurs in the course of transmission, so that there is no qualitative difference between transmission and preadolescent-mediated diffusion.

Sometimes, however, the distribution of a regular sound change across languages betrays its diffusion. Oceanic languages spoken on the islands and coasts of the St George's Channel separating New Ireland from New Britain provide a clear example. The languages all belong to the St George linkage, but it is easy to show (on the basis of free personal pronouns and of functors in noun phrases) that within it they form the three subgroups listed in (3).¹⁰

- (3) a. Patpatar (NI), Vinitiri [Minigir] (NB), Tolai (NB)
- b. Label (NI), Bilur [Birar] (NB)
- c. Kandas (NI), Ramoaaina (Duke of York Islands)

The firstnamed member of each group is spoken on the the New Ireland (NI) coast. Vinitiri is spoken in villages around Ataliklikun Bay on the north coast of the Gazelle Peninsula of New Britain which before modernization were quite isolated (Ross 1988: 258, 267, Van Der Mark 2007). These languages retain earlier *s as *s*. In the lastnamed member of each group, on the other hand, *s has become zero. The three *s*-less languages are geographically very close to each other,¹¹ closer than they are to their genealogically closest relatives, and *s*-deletion must have diffused from one of the languages, probably Tolai (with the largest speech community) to the other two. Its very regularity says that it was a preadolescent-mediated diffusion. Children in Bilur and Ramoaaina villages grew up bilingual in their heritage language and Tolai, and, as the change was proceeding in Tolai, they copied it in their heritage language. It is possible, too, that adults copied the change in order to identify with Tolai speakers, but its regularity can be attributed to

¹⁰ Square brackets enclose alternative language names.

¹¹ Tolai and Bilur are both spoken on the east coast of the Gazelle Peninsula of New Britain (NB), the small Bilur community being flanked on all sides by Tolai speakers. Ramoaaina is spoken on the Duke of York Islands, visible from the Tolai-speaking coast.

preadolescent learning. I suggest in §§5.2.2-4 that certain changes in the ‘Tsouic’ languages and Rukai are the outcomes of this kind of rule-copying.

I mentioned above that the probability of a sound change satisfying the ‘no independent innovation’ condition is not high. The possibility that a sound change may be diffused means that sound changes also fail to satisfy the ‘no copying’ condition. This is the reason why Sagart (2004) rejects Blust’s (1999) subgrouping of Formosan languages, which is based on phonological mergers.¹² However, whereas Sagart completely rejects sound change as a subgrouping indicator, I think that a sound change remains an indicator when it is a member of a coterminous set of innovations.

2.2.2.3 Changes in lexicon

Two kinds of lexical change are used in subgrouping arguments: replacements and idiosyncratic changes in form. Zorc (1986) argues for a Philippine subgroup on the basis of a list of replacement innovations and Blust (1978) for an Eastern Malayo-Polynesian subgroup partly on the basis of a list of replacement and formal innovations.

A replacement innovation occurs when one lexical form is replaced by another with the same meaning. A lexical replacement always comes from somewhere, either through copying or through semantic shift in an existing lexical item.

Ringe et al. (2002:99) find replacement lexical innovations to be ‘the least secure evidence for subgrouping’ because of the possibility that an apparent shared innovation is the result of parallel changes in two languages, i.e. the ‘no independent innovation’ is not fulfilled. They also point out (2002:69) that, because of the root-node limitation (§2.1), it is sometimes difficult to know which word is original and which the replacement, a point which is particularly relevant to Sagart’s subgrouping hypothesis (§4.3). Finally, they observe that it is sometimes difficult or impossible to know whether the ‘no copying’ condition is fulfilled, as a word copied from a closely related language is not easy to detect.

An idiosyncratic formal change is one which affects just a single word, like the ones in (1b). Because of its idiosyncrasy it is likely to satisfy the ‘no independent innovation’ condition, but there is no guarantee that it will satisfy the ‘no copying’ condition.

¹² The illustrative example Sagart chooses, the eighteenth-century spread of the uvular articulation of /r/ from Paris to other European cities speaking French, German, Dutch and Danish (Trudgill 1974), is unfortunate. It is a phonetic rather than a phonological change, and it is clearly an instance of adult-mediated diffusion, motivated by a desire to imitate the French court. The change is still in train in Montreal, where the fact that it is adult-mediated is attested by its incomplete application by some speakers (Sankoff & Blondeau 2007).

Obviously, lexical copies are a common signal of contact. Copies are recognizable if the source and recipient languages reflected the phonemes of the protolanguage differently at the time of transfer, but otherwise not.

On the basis of the discussion above the high-order subgrouping hypotheses mentioned in §1 can be arranged in order of evidential strength from strongest to weakest, as follows: Ross (2009) is based on a paradigmatic change that includes morphological innovations, Ho (1998) and Blust (1999) on phonological innovations, and Sagart (2004) on lexical innovations. However, the fact that certain innovations have a weaker evidentiary value does not mean that they can be dismissed. Either they reflect historical events which should be integrated into our history of Formosan languages, or they can be shown to be misinterpretations of the data. I outline the subgroupings and discuss their evidentiary basis below, first discussing the NAn hypothesis, then looking more briefly at Blust's, Ho's and Sagart's subgroupings.

3. The Nuclear Austronesian hypothesis

Ross (2009) proposes that there is a Nuclear Austronesian subgroup containing all Austronesian languages except Puyuma, Tsou and Rukai. The NAn subgroup is defined by a single, rather complex, innovation—the nominalization-into-verb (Nom-into-V) innovation described in §3.1—that is unlikely to have occurred independently in two or more languages or to have been copied from one language to another.

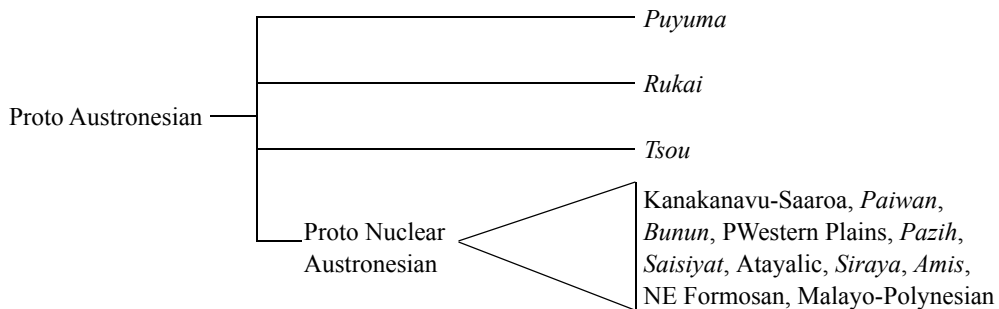


Figure 1: Primary Austronesian subgrouping according to the Nuclear Austronesian hypothesis

The resulting primary Austronesian subgrouping is shown in Figure 1. The figure acknowledges certain subgroups of Formosan languages (shown in roman, not italic, script) internal to NAn, which are the results of others' work and are discussed in §4.1 and §4.2. As Ross (2009) points out, the NAn hypothesis is compatible with Blust's (1999)

subgrouping of Formosan languages, with one exception: the Tsouic subgroup, which under the NAn hypothesis is broken into Tsou on one hand and Kanakanavu-Saaroa on the other (§5.2).

3.1 The nominalization-into-verb innovation

Table 1: A tentative reconstruction of Proto Austronesian verbal morphology, illustrated with reconstructed forms of PAn **kaen* ‘eat’

	ACTOR VOICE		UNDERGOER VOICE	
		Patient subject	Location subject	Circumstance subject
Realis nominalization	<i>*M-STEM</i> <i>*k<um>aen</i>	(no reconstruction)	<i>*STEM-an</i> <i>*kaen-an</i>	<i>*Sa-/Si-STEM</i> <i>*Sa-/Si-kaen</i>
Realis perfective nominalization	<i>*M-<in> STEM</i> <i>*k<um-in>aen</i>	<i>*<in> STEM</i> <i>*k<in>aen</i>	<i>*<in> STEM-an</i> <i>*k<in>aen-an</i>	(no reconstruction)
Irrealis nominalization	<i>*Ca-STEM</i> <i>*ka-kaen</i>	<i>*Ca-STEM-en</i> <i>*ka-kaen-en</i>	<i>*Ca-STEM-an</i> <i>*ka-kaen-an</i>	<i>*Sa-/Si-Ca-STEM</i> <i>*Sa-/Si-ka-kaen</i>
Realis	<i>*M-STEM</i> <i>*k<um>aen</i>	<i>*STEM-aw</i> <i>*kaen-aw</i>	<i>*STEM-ay</i> <i>*kaen-ay</i>	<i>*an-ay + STEM</i> <i>*an-ay kaen</i>
Optative/hortative	<i>*M-STEM-a</i> <i>*k<um>aen-a</i>			
Realis imperfective	<i>*M-Ca-STEM</i> <i>*k<um>a-kaen</i>	<i>*Ca-STEM-aw</i> <i>*ka-kaen-aw</i>	<i>*Ca-STEM-ay</i> <i>*ka-kaen-ay</i>	<i>*an-ay + Ca-STEM</i> <i>*an-ay ka-kaen</i>
Imperative	<i>*STEM</i> <i>*kaen</i>	<i>*STEM-u/-i (?)</i> <i>*kaen-u/-i</i>	<i>*STEM-i</i> <i>*kaen-i</i>	<i>*an-i + STEM</i> <i>*an-i kaen</i>
Dependent	<i>*M-STEM</i> <i>*k<um>aen</i>	<i>*STEM-a</i> <i>*kaen-a</i>		
Irrealis	<i>*Ca-STEM</i> <i>*ka-kaen</i>	(no reconstruction)	<i>*Ca-STEM-i</i> <i>*ka-kaen-i</i>	<i>*an-i + Ca-STEM</i> <i>*an-i ka-kaen</i>

Table 2: Proto Nuclear Austronesian verbal morphology, illustrated with reconstructed forms of PNAn **kaen* ‘eat’
(= Proto Austronesian verbal morphology as previously reconstructed)

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Realis (V/N)	<i>*M-STEM</i>	<i>*STEM-en</i>	<i>*STEM-an</i>	<i>Sa-/Si-STEM</i>
	<i>*k<um>aen</i>	<i>*kaen-en</i>	<i>*kaen-an</i>	<i>*Sa-/Si-kaen</i>
Realis perfective (V/N)	<i>*M-<in> STEM</i>	<i>*<in> STEM</i>	<i>*<in> STEM-an</i>	<i>*<in>Si-STEM</i>
	<i>*k<um>-in>aen</i>	<i>*k<in>aen</i>	<i>*k<in>aen-an</i>	<i>*S<in>i-kaen</i>
Realis imperfective (V/N)	<i>*M-Ca-STEM</i>			<i>*Sa-/Si-Ca-STEM</i>
	<i>*k<um>a-kaen</i>	<i>*Ca-STEM-en</i>	<i>*Ca-STEM-an</i>	<i>*Sa-/Si-ka-kaen</i>
Irrealis (V/N)	<i>*Ca-STEM</i>	<i>*ka-kaen-en</i>	<i>*ka-kaen-an</i>	<i>*Ca-STEM</i>
	<i>*ka-kaen</i>			<i>*ka-kaen</i>
Optative/hortative	<i>*M-STEM-a</i>	<i>*STEM-aw</i>	<i>*STEM-ay</i>	<i>*an-ay + STEM</i>
	<i>*k<um>aen-a</i>	<i>*kaen-aw</i>	<i>*kaen-ay</i>	<i>*an-ay kaen</i>
Imperative		<i>*STEM-u</i>		
	<i>*STEM</i>	<i>*kaen-u</i>	<i>*STEM-i</i>	<i>*an-i + STEM</i>
Dependent	<i>*kaen</i>	<i>*STEM-a</i>	<i>*kaen-i</i>	<i>*an-i kaen</i>
		<i>*kaen-a</i>		

Tables 1 and 2 show respectively the PAn and PNAn systems of verbal morphology reconstructed by Ross (2009), together with reconstructed forms of PAn/PNAn **kaen* (**[kaʔən]*) ‘eat’.¹³ In both tables **M-* represents a morphological unit with four exponents: the infix **<um>*, the replacement of stem-initial **p-* by **m-*, the prefix **ma-*, and zero. The difference between the PAn and PNAn systems was brought about by the Nom-into-V innovation. The PAn forms above the line in Table 1 are nominalizers. Above the line in Table 2, however, these forms double as both nominalizers and PNAn indicative (realis and irrealis) finite verbs, thereby displacing the indicative (realis) forms below the line in Table 1 from their indicative function. Hence the label ‘Nom-into-V’: what were once only nominalizations came in PNAn to function as both nominalizations and indicative verbs. The indicative forms below the line in Table 1 are henceforth

¹³ The PNAn reconstruction in Table 2 continues a line of reconstruction that started with Wolff (1973) and continued with Ross (1995a, 2002a, 2009).

labelled ‘first-generation forms’, the indicative forms above the line in Table 2 ‘second-generation forms’.

The early Austronesian Nom-into-V innovation was originally suggested by Andrew Pawley in lectures at the 1977 Summer Institute of the Linguistic Society of America, then elaborated by Starosta, Pawley & Reid (SPR) in a 1981 conference paper, published in abbreviated form in 1982, and finally published in full as Starosta, Pawley & Reid (2009). Meantime, Ross (1995a:767-768, 2002a:39-40) presented an internal reconstruction whereby the PAn affixes **M-...-a*, **-aw*, **-ay* (optative/hortatives in Table 1) were derived from a suffix **-a* plus **M-*, **-u* and **-i* respectively (imperatives and dependents in Table 1) and inferred that this was probably the backbone of the system before the Nom-into-V innovation occurred. This conclusion was based on the fact that **-a*, **-aw*, **-ay* and zero, **-u* and **-i* display the kind of paradigmatic patterning (in this case a set of suffixes) that one expects in verbal morphology—a patterning which the forms derived from nominalizations, namely **<in>*, **-en*, **-an* and **Sa-/Si-* (an infix, two suffixes and a prefix), do not display. The Puyuma data confirm this.

Ross (2009) was largely concerned with reconstructing the PAn and PNA systems. The reconstruction was provoked by the realization that Puyuma verbal morphology, described in detail by Teng (2008), more closely resembled SPR’s and Ross’s reconstructions of pre-PAn than of PAn.¹⁴ The morphemes reconstructed by SPR as erstwhile nominalizers are still nominalizers and nothing else in Puyuma. This implied either that Puyuma reflected a pre-PAn genealogical node, or that our view of PAn verbal morphology needed revision.¹⁵ This was really a choice between two sets of labels: (i) pre-PAn and PAn, or (ii) PAn and PNA. The first alternative was simply silly, as Tsou and Rukai also fail to reflect previously reconstructed PAn verbal morphology, and so Puyuma, Tsou and Rukai would have been cast out of the Austronesian fold. Hence the second pair of labels was chosen.

In 1995 I wrote, ‘The verb system of Puyuma, where earlier projective morphemes have ousted their neutral counterparts, also needs investigating’ (p.770). In other words, I reconstructed PAn much as I now reconstruct PNA, and inferred that Puyuma had simply lost the finite verbal affixes derived from nominalizations. I call this hypothesis, which I now think was wrong, the ‘loss hypothesis’. Loss hypotheses have also been proposed to account for the absence of evidence for the Nom-into-V innovation in Tsou

¹⁴ The outlines of Puyuma verbal morphology were already known from Tsuchida (1980) and Cauquelin (1991), but I did not recognize its implications until I read Teng’s more detailed account, especially of Puyuma nominalization.

¹⁵ Starosta (1996) came to a similar set of conclusions on the basis of Tsou and Rukai (he did not mention Puyuma), but I did not see this paper until after its republication in Zeitoun (ed., 2009).

(§5.1.6) and Rukai. The simplest explanation of these apparent facts is that these three languages represent first-order offshoots of PAN in which the innovation has not taken place (Ross 2009). Sagart (2010) revived a version of the loss hypothesis with regard to Puyuma, but Teng & Ross (2010) have shown that this is flawed.

The complexity of the Nom-into-V innovation lies in the replacement of a paradigm of first-generation finite verb morphemes by first-generation nominalizers that became second-generation finite verb morphemes. This wholesale morphological substitution is complex enough to make independent parallel innovation decidedly improbable, but there are three matters of shared detail that make it even more improbable. The first and second are the analogical geneses of the PNAn realis imperfective forms in **Ca-* and the patient-subject realis form **STEM-en*, the third the fact that the nominalizing affixes reanalyzed as PNAn finite verb forms were only a selection of the existing PAN nominalizers.

3.2 Realis imperfective forms in **Ca-*

The PNAn imperfectives are clearly reconstructable, all with **Ca-* reduplication, as the forms in (4) indicate.¹⁶

(4)	ACTOR		UNDERGOER VOICE	
	VOICE	patient subject	location subject	circumstance subject
PNAn	<i>*M-Ca-STEM</i>	<i>*Ca-STEM-en</i>	<i>*Ca-STEM-an</i>	<i>*sa/si-Ca-STEM</i>
Saaroa	<i>M-Ca-STEM</i>	<i>Ca-STEM-a</i>	<i>Ca-STEM-a[na]</i>	—
Saisiyat	<i>CV-M-STEM</i>	—	—	—
Pazih	<i>CV-STEM</i>	<i>CV-STEM-en</i>	<i>CV-STEM-an</i>	<i>sa-CV-STEM</i>
Thao	<i>Ca-M-STEM</i>	—	—	—
Bunun	<i>M-CV-STEM</i>	<i>CV-STEM-en</i>	<i>CV-STEM-an</i>	<i>si-CV-STEM</i>
Siraya	<i>M-Ca-STEM</i>	<i>Ca-STEM-en</i>

In the light of the Nom-to-V innovation one would expect these forms to have occurred as imperfective nominalizations in PAN, but the only PAN nominalizations reconstructable with **Ca-* are irrealis (Table 1), predictably becoming PNAn irrealis indicatives (Table 2).

The PNAn second-generation realis imperfectives with **Ca-* reduplication must have been formed by a combination of (i) analogy with the PAN first-generation realis imperfectives, also formed by **Ca-* reduplication but with first-generation suffixes, and (ii) functional extension of the patient- and location-subject irrealis indicatives (Ross 2009:307). This innovation of PNAn realis imperfective forms is unlikely to be the

¹⁶ The fact that **Ca-* is reflected as **CV-* in some NAn languages is handled in Ross (2009:298).

outcome of parallel independent innovation and speaks further for the claim that the Nom-to-V innovation occurred just once.

3.3 PAn *-en and *-an

It has been assumed (e.g. by SPR) that PAn *-en and *-an both formed nominalizations. There was, however, a decided mismatch between their functional loads. The PAn suffix *-an functioned both as a locative nominalizer and as a general nominalizer, as the structures in (5) show.

(5)	meaning	form	languages in which reflected
	general nominalizer	*STEM-an	Nanwang Puyuma, Tsou, Rukai, Paiwan, Thao, Pazih, Kavalan ¹⁷
	general perfective nominalizer	*⟨in⟩-STEM-an	Nanwang Puyuma, Rukai, Kanakanavu, Pazih, Saisiyat, Mayrinax Atayal, Paiwan, Amis, Kavalan
	general imperfective or irrealis nominalizer	*Ca-STEM-an	Nanwang Puyuma, Tsou, Paiwan, Thao ¹⁸
	location	*ta-STEM-an	Rukai, Tsou, Kanakanavu, Pazih, Amis
	circumstance	*Sa/Si-STEM-an	Rukai, Paiwan, Thao, Kavalan
	time	*ka-STEM-an	Puyuma, Rukai, Paiwan

The suffix *-en, on the other hand, is not reflected in Tsou or Rukai, and its only Puyuma reflex is in *Ca-STEM-en* ‘patient-subject irrealis nominalization’ in the Tamalakaw (Tsuchida 1980:203, 208) and Katripul (Stacy Teng, p.c.) dialects.

Where then did the ubiquitously reflected PNA **STEM-en* ‘patient-subject realis’ verb form come from? It is reflected in Kanakanavu *STEM-ene*, Saaroa *STEM-a*, Saisiyat, Pazih, Paiwan and Siraya *STEM-en*, Mayrinax Atayal, Seediq and Ishbukun Bunun *STEM-un*, Thao *STEM-in*, and also Central Amis *STEM-en* ‘patient-subject future’. Since PNA second-generation forms were derived from PAn nominalizations, we would expect the parent of PNA **STEM-en* to have been PAn **STEM-en* ‘patient-subject realis nominalization’—but such a PAn form is not reconstructable. Instead, at some pre-PNA

¹⁷ **STEM-an* is reflected exclusively as a location nominalizer in Saisiyat and Amis.

¹⁸ **Ca-STEM-an* is reflected exclusively as a location nominalizer in Pazih, Saisiyat and Mayrinax Atayal, and as a circumstance nominalizer in Amis.

stage the forms in (6) apparently existed, and speakers filled the gap in the paradigm with *STEM-en.

(6) <i>finite verb and nominalization</i>		
	<i>patient-</i>	<i>location-</i>
	<i>subject</i>	<i>subject</i>
realis	—	*STEM-an
irrealis	*Ca-STEM-en	*Ca-STEM-an

In other words, PNAn *STEM-en ‘patient-subject realis’ was created by analogy with (pre-)PNAn *Ca-STEM-en ‘patient-subject irrealis’, which is attested in Pazih, Mayrinax Atayal and Central Amis. Only Pazih and Paiwan reflect *STEM-en as a nominalization, suggesting that PNAn *STEM-en was initially a finite verb form and that the Paiwan and Pazih nominalizations are analogical back-formations.

Thus *-en must have had a very low functional load in PAn but acquired a much more central function as the main patient-subject morpheme in PNAn.¹⁹ Again, because this change affects a particular form, it is unlikely to have occurred more than once in the early history of Austronesian and is additional evidence for PNAn.

3.4 Selection of PAn nominalizations for reanalysis

Another dimension of the Nom-to-V innovation is that only a subset of available PAn nominalizations was reanalyzed as PNAn finite verbs, and NAn languages agree on the forms that were selected for this reanalysis. In addition to the nominalizers above the line in Table 1, several other nominalized forms occurred in PAn.

(7)	meaning	form	languages in which reflected
	actor	*ta-STEM	Rukai, Tsou, Pazih, Sqliq Atayal
	location	*ta-STEM-an	Rukai, Tsou, Kanakanavu, Pazih, Amis
	circumstance	*Sa/Si-STEM-an	Rukai, Paiwan, Thao, Kavalan
	circumstance	*Ca-STEM	Rukai (?), Thao, Pazih, Saisiyat
	time	*ka-STEM-an	Puyuma, Rukai, Paiwan

¹⁹ An alternate interpretation is possible, namely that Puyuma -en was copied from Paiwan (we know that contact occurred; Blust 1999:47-51), and that *-en did not occur in PAn but was entirely a PNAn innovation.

Why, for example, was PAn *⟨um⟩STEM selected for the actor voice rather than *ta-STEM? There is no obvious explanation, but one possibility should be aired, despite its improbability.

Comrie & Thompson (1985:349) make a primary typological cut between event nominalizations and argument nominalizations. PNAn verb forms, which encapsulate the semantic role of their subject, are derived from argument nominalizations. Comrie & Thompson observe that across the world's languages argument nominalizations typically do not retain features such as the valency, voice, mood and aspect of the verbs from which they are derived. But they note that Tagalog is exceptional in this respect (p.352): its argument nominalizations do retain verbal features, and this is typical of Formosan and Philippine languages. What if PAn made a distinction between argument nominalizations that retained more verbal features and argument nominalizations that retained fewer or no verbal features? And what if PAn *⟨um⟩STEM was a nominalization of the former kind and PAn *ta-STEM of the latter? In this case, it would be easy to see why *⟨um⟩STEM was reanalyzed as a verb, but not *ta-STEM. However, Comrie & Thompson (1985) find no language where such a morphological distinction is made, and it is poor practice to reconstruct a morphological distinction that is not known in a modern language. It is thus quite possible that the PNAn selection of certain nominalized forms for reanalysis was fortuitous, and yet another piece of evidence that the Nom-into-V innovation is very unlikely to have occurred independently in more than one language.

3.5 Assessing the evidence

The NAn subgroup is founded on a single innovation, the Nom-into-V innovation: there is no bundle of coterminous innovations here. However, it is an innovation that fully satisfies the 'no copying' and 'no independent innovation' conditions. Not only is it unlikely that a morphosyntactic innovation on this scale would be copied, but the literature on contact-induced morphosyntactic change suggests strongly that it is almost always only syntax that is copied, not affixal forms (§2.2.2.1). As it is, NAn languages reflect as verbal affixes the same **subset** of PAn nominalizing affixes (§3.4), strongly supporting the claim that the Nom-into-V innovation occurred just once—in PNAn.

Thanks to the absence of external evidence (due to the root-node limitation, §2.1), the only direct evidence for reconstructing the PAn realis and irrealis finite verb forms in Table 1 is the presence of these forms with these functions in Puyuma. One language is arguably not great evidence, but there are numerous instances in historical linguistics where data from a language on the periphery of a family have lent significant evidence for its reconstruction: the majority is not necessarily right. In this instance, there is also indirect support in the shape of concurrence of Puyuma with the internal reconstruction discussed in §3.1.

No phonological innovations define NAn, but there is a lack of serious phonological evidence for **any** high-order subgroup (§4.1) other than that for Malayo-Polynesian referred to in (1). It would be quite easy to produce a list of hundreds of apparent PNAn lexical innovations, that is, items not reflected in Puyuma, Tsou or Rukai. But to be sure that the items on that list were actually PNAn innovations, one would have to be reasonably certain that they do not occur in the three languages. However, it is questionable whether the available lexical materials for these languages allow us to be so certain. We are again faced with the root-node limitation. A single lexical cognate in a sister language of PAn would be enough to tell us that a lexical item occurred in PAn, even if it has since been lost in Puyuma, Tsou or Rukai—but we don't have such evidence. There are only two convincing ways in which lexical evidence could be offered for PNAn. The first would consist of idiosyncratic changes in the forms of PNAn lexical items relative to the PAn form reflected in Puyuma, Tsou and Rukai (§2.2.2.3). I have not found such forms. The other would come from the distributions of cognate sets across the Formosan languages and Malayo-Polynesian. If a distribution whereby cognates sets were restricted to NAn languages occurred significantly more frequently than other distributions (where 'significantly' is used in its statistical sense), then we would have a lexical case for NAn. But again one would need better lexical materials right across the Formosan languages than are currently available.

4. Other high-order subgroupings of Austronesian and the evidentiary value of innovations

4.1 Blust (1999)

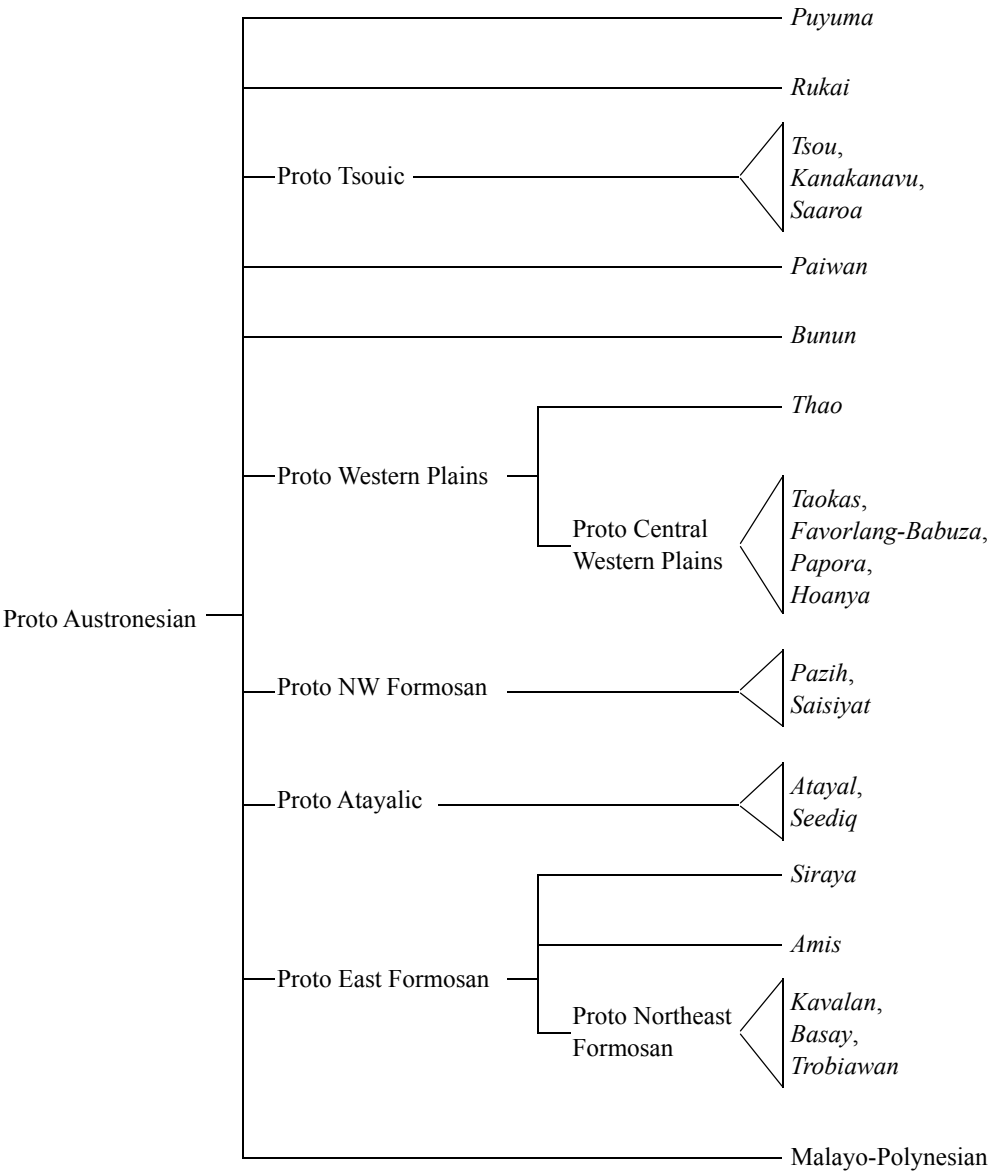


Figure 2: Primary Austronesian subgrouping according to Blust (1999)

Blust's subgrouping (Figure 2) is based almost entirely on phonological innovations (1999:44-45). Table 3 shows the PAn consonants. Those printed in italics are not part of Blust's paradigm but are argued for by Ross (2011). Table 4 presents the sound correspondences on which Table 3 is based. The putative phonetic values are mine. The row labelled 'step 1' in Table 4 represents a fairly conventional set of inferences about the phonetic values of PAn phonemes, while the row labelled 'step 2' represents a set of inferences based on the assumption that Paiwan is phonetically more conservative than all other Austronesian languages. The two sets of inferences are discussed by Ross (2011). In Table 3 the format '[x], [y]' represents these alternative ascriptions of phonetic values, where [x] represents 'step 1', [y] 'step 2'.

Table 3: PAn consonant paradigm with putative phonetic values
(italicized protophonemes are proposed by Ross 2011; the format [x], [y] represents alternative ascriptions of phonetic values discussed by Ross 2011)

	Bilabial	Dental and inter- dental	Alveolar	Laminal post- alveolar	Palatal	Velar	Uvular and pharyngeal
stop vl	p		t [t], [c]		s [c], [t]	k	q
stop vd	b [b, β]		d [d], [ɟ]	<i>D</i> [ɖ]	<i>Z</i> [ʈ], [ɟʰ]	g	
affricate vl			C [ts]				
affricate vd			z [dz]				
fricative vl		θ [θ]	x [s]	S [ʃ, ʃ]			h [h]
fricative vd					j [j, ʲ], [z]		
nasal	m		n		ñ [ɲ]	ŋ	
lateral			L [l, ɭ], [ɭ, ɭ]	l [l, ɭ]			
trill							R [r] or [ʀ]

Table 4: Formosan and PMP consonant correspondences

Stops and affricates													
Position	Bilabial			Alveolar		Postalveolar			Palatal	Velar	Uvular	Glides	
	</												

Notes: In phonetic representations dialect alternants are separated by a slash (/), other alternants by a comma.

1. Blust and I now write earlier *Z as *z, but in my orthography it contrasts with *Z, a phoneme not recognized in the conventional orthography.

Table 4 (continued): Formosan and PMP consonant correspondences

Position	Fricatives			Nasals			Liquids				
	Interdental	Alveolar	Postalv.	Pal./Vel.	Glottal	Bilabial	Alveolar	Palatal	Velar	Alveolar	Postalv
PAn (Tsouchida)	*θ	*S ₂	*S ₁	*j	*H	*m	*n	*ñ	*ŋ	*L, *N	*I
PAn (Blust)	*s	*S	*S	*j	*h	*m	*n	*ñ	*ŋ	*N	*I
PAn (Ross)	*θ	*X	*S	*j-j	*h	*m	*n	*ñ	*ŋ	*L	*I
Puyuma	s	θ	θ	d	θ	m	n	l	ŋ	l	l [l]
P Rukai	θ-θ-s	s	S	g, θ	θ	m	n	...	ŋ	l	l [l]
Tsou	s [s, ʃ]	θ	s	θ	θ	m	n	n	ŋ	h, k/_f, s, h	r [r/ʃ]
Kanavu	s [c, ʃ]	θ	s-s/θ-s	l [r], θ	θ	m	n	ŋ	ŋ	n	l [r]
Saaroa	s [s, ʃ]	θ	s/θ-θ-s/θ	l [ʃ], θ	θ-θ-?	m	n	ʔ [ʃ]	ŋ	ʔ [ʃ]	r
Paiwan	t	s	s	d	-θ	m	n	ʔ [ʃ]	ŋ	ʔ [ʃ]	θ, r
Bunun	c [tʃ]	s-s, θ-s	s	θ	h [χ]	m	n	n	ŋ	n	l
Thao	t	θ	ʃ (sh)	θ (z)	-?	m	n	θ (z)	n	θ (z)	ʔ (th)
Taokas	t	θ	ʃ	t, θ		m	n	...	n	t	h
Fav-Babuza	t	θ	s	d		m	n	...	n	s, l	θ
Papora	t	θ	s	d, θ		m	n	...	n	l	l (?)
Hoanya	t	θ	s	dz		m	n	...	ŋ	s	h
Pazih	z-z-t	h [h]	s	z, -t	h [h]	m	n	...	ŋ	l	r [r-ʔ]
Saisiyat	h	h	ʃ	z [ð]	h	m	n	l	ŋ	l	l [l]
P Atayal	h [h]	h [h]	s [ʃ]	j [j]	h [h]	m	n	l [ʃ]	ŋ	l	r [r-ʔ]
Siraya	s	θ	x-θ-x, θ, ŋ	n [ŋ]	h-θ [h]	m	n	l	ŋ	l	g [g~ʃ]
P Amis	c [ts]	s	s	n		m	n [n]	d [d, ð, ʃ, ʔ]	ŋ	d [d, ð, ʃ, ʔ]	l [r]
Basay	ts	ts	s	n		m	n	...	ŋ	n	l, θ
Trobiawan	ts	ts	s	n		m	n	...	ŋ	n	l
Kavalan	s	s	s	n		m	n	n	ŋ	n	r [r]
PMP	*s	*h	*h	*j		*m	*n	*ñ	*ŋ	*l-n-n	*I
											*r, R

Those of Blust's innovations that are reflected in more than one language are set out in Table 5.²⁰ The table format shows which languages reflect which innovations. A bullet (•) indicates a categorical innovation, a plus sign (+) a partial innovation, i.e. one that does not occur in every candidate lexical item, and a question mark (?) indicates that one of the phonemes involved in the merger is not attested in the data.

Table 5: Shared phonological innovations noted by Blust (1999) in Formosan languages

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

²⁰ Kulon is omitted, as Blust says that the data are really too sparse to be reliable.

Note: Mergers shown in italics are modifications of or additions to Blust's mergers, based on Ross (2011).

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

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

All these innovations are mergers, but Blust includes among them mergers with zero, i.e. deletions. I show these as deletions, placing them with other lenitions at the top of the table. I have dismantled mergers involving more than two phonemes (including zero). For example, Taokas *p/k/j/∅ is shown as three deletions, *p > ∅, *k > ∅ and *j > ∅, and Kavalan *j/n/L as three mergers, *j/*n, *j/*L and *n/*L,²¹ since the three changes are not necessarily connected. This makes it easier to compare innovations across languages and to identify shared innovations.²² I have modified Blust's analysis by adding the PAn phonemes *D, *Z, *x and *θ posited by Ross (2011).²³ Innovations that involve them are shown in italics, but in practice they make no difference to Blust's subgrouping.

Table 5 reveals that **no** group of Formosan languages is defined by a bundle of coterminous innovations. Instead, what we find are numerous overlapping innovations, as we would expect from a collection of languages that have remained in contact with each other since their common genesis. The PMP situation is different, as each blob represents an innovation reflected across Malayo-Polynesian languages and, as noted in (1a-i) and (1a-ii), the *S/*h merger and the mergers shown as *L > *l-n-n in the table are coterminous, along with several more detailed phonological innovations.²⁴

Blust uses mergers and deletions because they are irreversible, and therefore 'particularly powerful evidence of exclusively shared history', but adds the qualification that 'the subgrouping value of a merger ... depends on its distinctiveness' (1999:42). In the terms used here 'distinctiveness' means satisfying the 'no independent innovation' condition (1999:45). As Figure 2 shows, he identifies ten first-order subgroups of Austronesian languages: nine in Taiwan and Malayo-Polynesian (§2.2.1) outside Taiwan. Four of the nine Formosan subgroups have only a single member (Puyuma, Rukai, Paiwan and Bunun). The other five are Tsouic, Western Plains, Northwest Formosan, Atayalic and East Formosan. Blust (1999:52) simply says that Tsouic has been established by Tsuchida (1976): I discuss the putative Tsouic subgroup in detail in §5.2. I will discuss each of the other four groups briefly here.

²¹ Blust writes *N where I prefer *L, as this was probably a dental lateral (Ross 2011).

²² By implication, Blust does the same, as he identifies *j/*n as the major innovation diagnostic of the East Formosan subgroup.

²³ PAn *Z is the *d₂ of Dahl (1976) and Ross (1992) (distinct from Blust's and my *z, formerly written as *Z); *D is *d₃. PAn *x is Tsuchida's (1976) *S₂, Li's (1985b) *ʃ; PAn *θ is Tsuchida's and Li's *θ.

²⁴ These are *S metathesis (Blust 1993), PAn *-eS > PMP *-ah (Blust 1995a:623-624) and PAn *-RVj > PMP *-IVj (Blust 2001:152-153).

The Western Plains group consists of the Central Western Plains group (Taokas, Favorlang-Babuza, Papora and Hoanya) established by Tsuchida (1982) on the basis of wordlists from these now extinct languages, and Thao, which shares certain lexical innovations with the Central Western Plains languages (Blust 1996). One phonological merger, of non-final *s and *t,²⁵ is common to all five languages (Blust 1999:44) and found nowhere else in Taiwan. Two more mergers mentioned by Blust are partial. The word-final diphthongs *-ay and *-aw lose their final glide in the Central Western Plains languages, but the innovation does not apply to all candidate lexical items in Thao. The merger of *n/*ŋ is incomplete in Hoanya (PAN *laCej ‘vegetables’ > Hoanya *lasen* ‘vegetables’, but PAN *Canjis ‘weep’ > Hoanya *s(am)ani*). The Western Plains group is thus supported by the *s/*t merger and lexical innovations. The partial innovations testify to contact rather than inheritance.

The Northwest Formosan group is the most weakly supported of the four discussed here, and Blust (1999:52) says that ‘the evidence ... is not strong’. The group contains just two languages, Saisiyat and Pazih, which share two innovations. They are not, however, coterminous. The *C/*s merger is shared by Taokas, Papora and Hoanya. The second innovation, whereby *q became ? in Saisiyat and was lost in Pazih (presumably after becoming ?), is also reflected in a number of Formosan languages (Table 5).

Blust (1999:46) labels the Atayalic group (Atayal and Seediq) ‘self-evident’ and says it ‘has been adequately demonstrated in e.g., Li (1981)’. This seems to contradict his observation under ‘significant mergers’—‘virtually nothing, as the mergers that occur are characteristic of many other languages’ (1999:44)—but I take it that he considers Li to have demonstrated the lexical unity of Atayal. Li (1981:275) provides only one Proto Atayalic innovation which satisfies the ‘no independent innovation’ condition, namely a conditioned split whereby PAN *e becomes Proto Atayalic *u in final syllables and *e elsewhere.²⁶ This split is found in no other Formosan language. Harvey (1982:89) noted that the Atayalic languages uniquely share the first person singular genitive enclitic =*mu*.²⁷ But for these two innovations, one could argue that the Atayalic dialects form a network that has subsisted since early Austronesian times and that the similarities of its dialects are due to innovations that have spread through the network, an assertion supported by the fact that the boundary between Atayal and Seediq is difficult to draw

²⁵ In Thao also in final position.

²⁶ Li (1981:274-275) lists other phonemic splits, but they are not supported by the data available to me.

²⁷ For the evidence, see Ross (2006:549). Harvey also claims that Atayalic dialects uniquely share the first and second person singular locative pronouns reflecting Proto Atayalic *ke-nan and *su-nan, but they have cognates in other Formosan languages (Ross 2006:536).

(Li 1985a). These two innovations indicate, however, that in their earliest history the Atayalic dialects enjoyed a short period of unity.

Perhaps the most important finding of Blust (1999) is the East Formosan subgroup (Basay-Trobiawan; Kavalan, Amis and Siraya), but even here there are no coterminous innovations. The merger of *j and *n is exclusively reflected in these languages, but the merger of *t and *C is also found in Bunun and PMP, and Blust (1999:46) posits three independent *t/*C mergers, one in East Formosan, one in Bunun and one in PMP.²⁸ He argues that the merger of *j and *n is significant because it is highly unusual, occurring nowhere else in Austronesian. The principle to which Blust appeals, that rare sound changes are more likely to fulfil the ‘no independent innovation’ than common ones, is fine, but it raises the question, Did the merger truly entail a rare sound change? What were the phonetic values of the reflexes of *j and *n in an ancestor of the languages grouped as East Formosan? Blust (2009:573) thinks that the phonetic value of PAN *j was [g^j]; Ross (2011) takes it to have been a voiced fricative, either [j, ɣ] or [z], depending on one’s assumptions about the sound changes reflected in Formosan languages (the root-node limitation of §2.1 prevents certainty). However, in a number of Formosan languages the reflex of *j is an alveolar (Puyuma, Paiwan, Favorlang-Babuza, Papora *d*, Hoanya *dz*, Taokas *t*, Pazih *z*, Kanakanavu *l* [r]) and Saaroa *l* is a lateral fricative. The merger of [l] with [n] is not uncommon in Austronesian languages, and in Basay-Trobiawan and Kavalan PAN *L (probably [l] or [ɭ]) has also merged with *n ([n]). It is thus a reasonable possibility that in one or more immediate ancestors of the East Formosan languages PAN *j had become [l] before it merged with *n. If so, then we are not dealing with such an unusual sound change and it is possible that independent parallel innovation occurred not only in the case of *t/*C but also in the case of *j/*n.²⁹

Where does this discussion lead us? There is respectable if weak evidence for Western Plains and Atayalic, questionable evidence for East Formosan, and almost no evidence for Northwest Formosan. The phonological evidence in Table 5 points rather

²⁸ Blust mentions idiosyncratic changes in the forms of certain lexical items, but they are not coterminous with the *j/*n merger.

²⁹ An anonymous reviewer writes, ‘A merger that is unique in a language family... forces anyone who dismisses it as subgrouping evidence to explain why the same merger did not occur elsewhere. In other words, the subgrouping value of this merger is not in any way based on assumed radical differences in the phonetic properties of *j and *n, but only on the uniqueness of the merger.’ I disagree: it is the phonetic properties of phonemes that determine the probability of a sound change and thereby its subgrouping value. The fact that the *j/*n merger did not occur in other Formosan or Malayo-Polynesian languages may simply reflect the fact that sound changes are probabilistic, and *j instead underwent other changes. The outcomes of these changes were evidently values that blocked the possibility of a merger with the reflex of *n.

strongly to the gradual diversification of a dialect network into the present-day Formosan languages, with a brief period of shared unity in each of the earliest histories of Western Plains and Atayalic. It is not clear that the East Formosan languages enjoyed such unity, but a glance at Table 5 shows several coterminous innovations defining the Northeast Formosan grouping of Trobiawan-Basay and Kavalan.

4.2 Ho (1998)

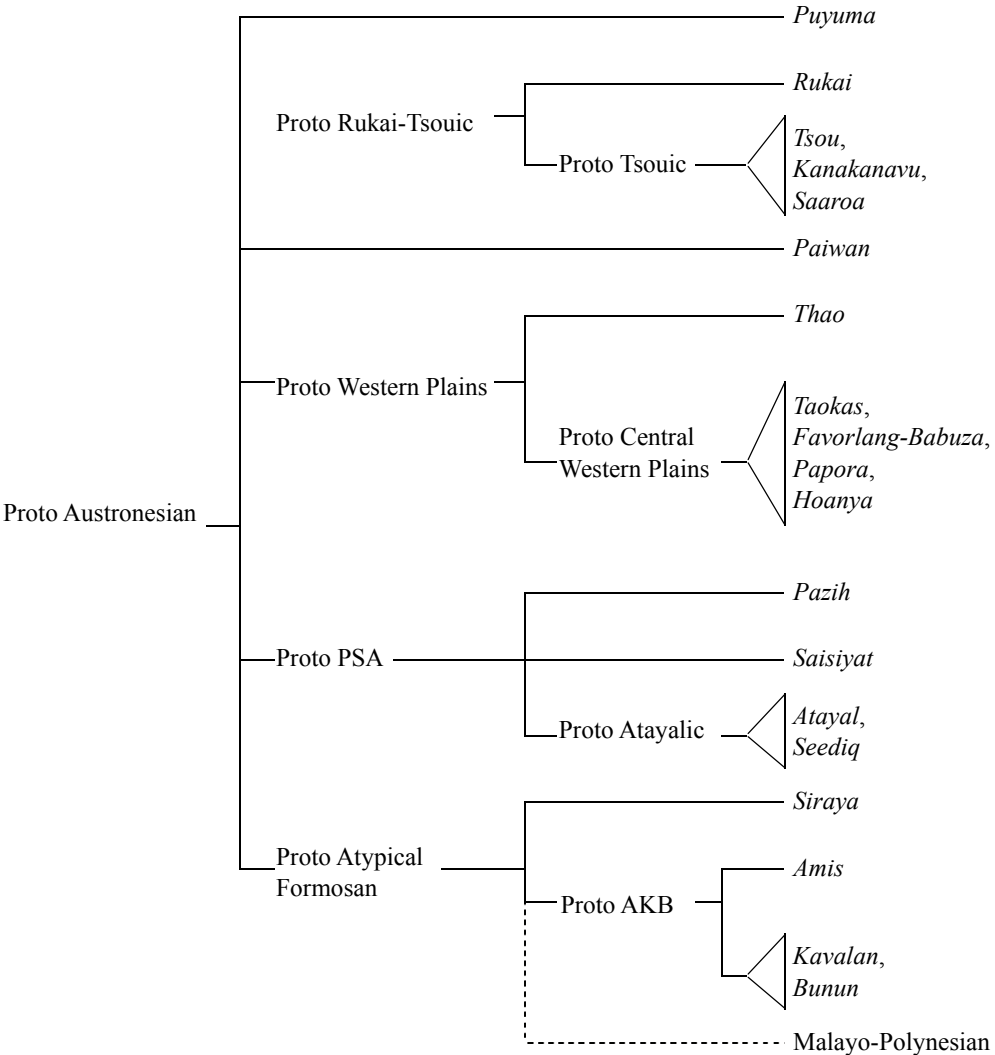


Figure 3: Primary Austronesian subgrouping according to Ho (1998)

Ho's subgrouping (Figure 3) is, like Blust's, based on phonological innovations, which he shows in a table (1998:145-146) that inspired my Table 5, replicating his format. Figure 3 incorporates certain liberties of my own. None of the nodes in Ho's (1998:158) figure are labelled, and I have named them in order to refer to them. In Ho's figure Pazih and Atayal are grouped, i.e. together they form a branch coordinate with Saisiyat beneath PSA, but his table shows no innovation defining a Pazih/Atayal group, so I have not included it in Figure 3. Finally, Ho does not discuss the position of PMP, but, since his Atypical Formosan group is defined by the **t/*C* merger, I have placed PMP there, indicating the addition by dashed lines.

Comparing Figures 2 and 3, we see both commonalities and differences between Blust's and Ho's subgroupings. Puyuma and Paiwan³⁰ form single-member primary subgroups in both. Both recognize Tsouic, Western Plains (with the same internal subgrouping) and Atayalic, but Ho also has a Rukai-Tsouic subgroup. Both recognize a grouping that includes Pazih and Saisiyat, but Ho combines it with Atayalic to form PSA (Pazih-Saisiyat-Atayalic). Both recognize a subgroup that includes Siraya, Amis and Kavalan,³¹ but Ho places Bunun (for Blust a single-language primary subgroup) in a low-order group with Kavalan.

Tsouic is discussed in detail in §5.2 and Rukai-Tsouic briefly in §5.3. Western Plains and Atayalic were discussed in §4.1. To the innovations cited by Blust in favor of Central Western Plains (Taokas, Favorlang-Babuza, Papora and Hoanya), Ho adds deletion of PAn **k*. The number of potential cases is few, but Hoanya *sikan* 'fish' (< PAn **Sikan*) suggests that deletion is incomplete in Hoanya and that this is therefore not an inherited shared innovation.

Ho's Pazih-Saisiyat-Atayalic group is based on the merger of PAn **d* with **D*. This is a phonemic distinction that Blust does not recognize, as he ascribes the reflexes of both to **D*. Ross (2011) accepts the distinction, but finds that it is only weakly supported. As Table 5 shows, according to Ross (2011) the **d/*D* merger is also reflected in Rukai, Tsou, Saaroa, Bunun, Siraya and Kavalan, and Ho's table of sound correspondences (1998:163) bears this out. It probably had also occurred in Kanakanavu (§5.2.5.2). The **d/*D* merger thus does not meet the 'no independent innovation' condition.

The basis for Ho's Atypical Formosan subgroup (Bunun, Siraya, Amis and Kavalan) is the merger PAn **C* with **t*. Since the merger is complete in every language that displays it, there is no reason to attribute it to copying. Whether it meets the 'no

³⁰ Ho's reason for placing Puyuma in a single-member subgroup is that it is otherwise difficult to determine its position because of contact with its neighbors. He considers Paiwan to be phonologically the most conservative Formosan language, an observation confirmed by Table 5.

³¹ The fact that Ho does not include the extinct languages Trobiawan and Basay is irrelevant, as they are so clearly closely related to Kavalan (Li 2001).

independent innovation' condition or not is a difficult to say. If the phonetic values of *C and *t were *[ts] and *[t] respectively, then the answer is perhaps 'yes', but if they were *[ts] and *[c] (giving priority to their Paiwan values), then the answer is probably 'no'. Blust's (1999:45) argument is that the *j/*n merger that defines his East Formosan subgroup (essentially Ho's Atypical Formosan minus Bunun) is a much stronger piece of evidence that the *t/*C merger, and so he assumes that the East Formosan *t/*C merger is coterminous with it, and that the *t/*C merger occurred independently in Proto East Formosan, in Bunun and in PMP. I suggested in §4.1 that the *j/*n merger is not as strong as it appears, and I suggest now that the only objection to treating the *t/*C merger as defining a subgroup is that no other innovation is coterminous with it.

Ho divides his Atypical Formosan first into Siraya and Bunun-Amis-Kavalan. The latter is based on the merger of PAn *S and *x (Ho's *S and *ʃ, Tsuchida's *S₁ and *S₂) as *s*. Blust treats the correspondence sets reflecting these two protophonemes both as reflexes of *S, but the difference between the two correspondence sets is sharp enough to justify to justify the reconstruction of two protophonemes. The difficulty here is that *x is reflected in only a small number of lexical items (see Ross 2011), and the Bunun reflexes of two etyma with medial *x, *kawi?* 'wooden' < PAn *kaxiw 'tree', and Bun *huuŋ* 'mushroom' < PAn *quzun, lack expected -s- (cf Bunun *ma-tusul* < PAn *CuxuR 'thread (needle)', Bunun *ma-nisbis* < PAn *Lixepis 'thin'). The implication is that either the zero or the -s- reflex entered Bunun through contact. Since the Bunun initial and final reflex of *x is *s*, it is more likely that the inherited Bunun reflex was -s-, and that Bunun therefore shared in the *S/*x merger, as Ho claims. Nevertheless, this is tenuous evidence on which to base a subgroup.

Ho's Bunun-Amis-Kavalan subgroup is divided into Amis and Bunun-Kavalan. In addition to the innovations already mentioned, Bunun and Kavalan share the merger of PAn *s and *θ, the merger of PAn *d, *z, *D and *Z, and the merger of *n, *ñ and *L. The first two mergers happen in several other Formosan languages and fail to satisfy the 'no independent innovation' condition. The number of PAn etyma containing *ñ is so low that we cannot be sure how many languages reflect the *ñ/*L merger, but Puyuma, Saaroa, Paiwan, Thao, Saisiyat, Atayalic, Siraya and Amis also do so, so it too has no subgrouping value. This leaves the *n/*L merger, which occurs in Saaroa, Bunun and Kavalan. Since Saaroa forms an obvious subgroup with Kanakanavu, we can be reasonably certain that the innovation occurred independently in Saaroa. This means we cannot be sure whether it also occurred independently in Bunun and Kavalan. That is, the evidence for a Bunun-Kavalan subgroup is vanishingly weak.

In summary, as I said at the end of the discussion of Blust's subgrouping, there is respectable if weak evidence for Western Plains and Atayalic. There is almost no evidence for Pazih-Saisiyat-Atayalic. There is respectable, if debatable, evidence for Atypical

Formosan, and tenuous evidence for an Amis-Kavalan-Bunun subgroup within it, but no serious evidence for the pair Bunun-Kavalan.

4.3 Sagart (2004)

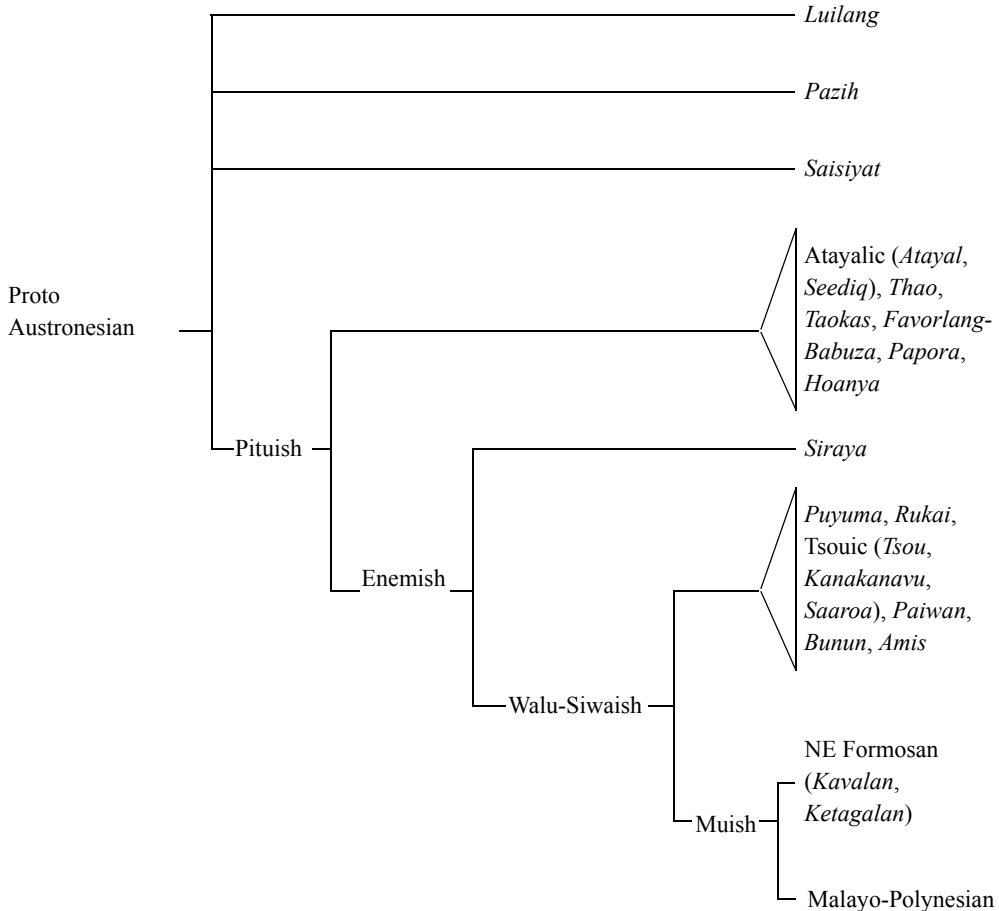


Figure 4: Primary Austronesian subgrouping according to Sagart (2004)

Sagart's subgrouping (Figure 4) is radically different from Blust (1999), Ho (1998) and Ross (2009), because the innovations it is based on are mainly idiosyncratic lexical innovations, and these lead to very different results from the phonological innovations employed by Ho and Blust or the complex morphosyntactic innovation described by Ross. Sagart's position on the evaluation of innovations is discussed in §2.2.2.2.

The argument on which Sagart's subgrouping is based begins with the observation (2004:413) that:

throughout Taiwan, a reflex of *puluq ‘10’ implies the presence of a reflex of *Siwa ‘9’, which implies the presence of *walu ‘8’, which implies the presence of *enem ‘6’, which implies the presence of *lima ‘5’, which implies the presence of *pitu ‘7’, while the reverse implications do not hold.

Table 6: Implicational hierarchy of the numerals 5-10 in Formosan languages and PMP, after Sagart (2004:414)

		‘7’	‘5’	‘6’	‘8’	‘9’	‘10’
Pituish	Luilang	in-nai	[na]lup	[na]tsulup	patulu-nai	satulu-nai	isit
	Saisiyat	saivuseaha	rrasu	saivusa	makaspat	ra:ha	ranpon
	Pazih	xasebidusa	xasep	xasebuza	xaseb[i,a]turu	xasebisupat	isit
	Favorlang	naito	achab	nataap	maaspat	tannacho	zchiett
	Taokas	yweto	hasap	tahap	mahalpat	tanasu	(ta)isid
	Hoanya	pito	Lima	<i>(mi)nun</i>	<i>(mi)alu</i>	<i>(a)sia</i>	<i>(miata)isi</i>
	Papora	pitu	nema	<i>(ne)nom</i>	<i>mahal</i>	<i>(me)siya</i>	<i>(me)tsi</i>
	Atayal	pitu?	imagal	czui?	spat	qeru	lpuu
	Seediq	pito	lima	mataro	maspat	maḡali	maxal
Enemish	Thao	pitu	rima	katuru, makalhturuturu	hashpat, maka(lh)shpashpat	tanacu	maqcin
	Siraya	pittu	rima	nəm	kuixpa	matuda	saat kittian
	Walun-Siwaish	pitu	eimo	nómə	vóeu	sío	máskə
	Saaroa	(k)upito	(k)ulima	(k)ənəmə	(k)ualo	(k)usia	(ku)ma:ɬə
	Kanavu	pitu	rima	nəm	(h)a:ru	si:ya	ma:nə
	Bunun	pitu?	hima?	nuum	vau?	siva?	masʔan
	Rukai	pitu	Lima	eneme	vaLu	baḡatə	maeale
	Paiwan	pitju	lima	enem, unem	alu	siva	puluq
	Puyuma	pitu	Lima	nem	waLu	iwa	puLu
Muish	Amis	pitu	lima	?enem	falu	siwa	polo
	Kavalan	pitu	rima	?nem	waru	siwa	betin
	Basay	pitu	tsjima	anəm	wasu	siwa	labatan
	PMP	*pitu	*lima	*enem	*walu	*siwa	*puluq

Table 6 (Sagart’s Table 1) displays the numeral forms on which this observation is based together with the resulting implicational hierarchy marked in grey. I have shifted Hoanya and Papora to place them with Taokas and Favorlang-Babuza, i.e. in the Central Western Plains group recognized by Tsuchida (1982), Blust (1999), Ho (1998) and Sagart. Sagart infers that the forms in italics are copied from Kakanavu or Saaroa, as they reflect *w as zero.³²

³² On the restricted available data it is difficult to know whether the Central Western Plains languages otherwise retained PAn *w. I know of just one example that attests retention (and none that attest certain loss): PAn *zawaC ‘walk’ > Hoanya dz<am>awat.

Because of the root-node limitation Sagart writes that the distribution of forms in Table 6 allows two hypotheses.

- A. *puluq etc did not occur in PAn, but are post-PAn lexical replacement innovations that define subgroups.
- B. *puluq etc are PAn forms that have been lost in certain Formosan languages.

Sagart argues for hypothesis A, and, as Figure 4 and Table 6 show,³³ he names his subgroups—with the exception of Muish—after the numeral forms that define them. He argues for A because (i) he believes that *pitu, *walu and *Siwa can each be derived from reconstructable PAn forms; and (ii) he thinks an implicational series of lexical replacements defining a nested subgrouping is more likely than an implicational series of lexical losses.

I will argue that the case for hypothesis A is weak and the case for hypothesis B considerably stronger.

4.3.1 Hypothesis A

Sagart's derivations of *pitu, *walu and *Siwa are shown in (8). The bolded segments in the putative PAn forms on the left are those from which the forms on the right are said to be derived.

- | | | | | |
|-----|------------------------|-----|---|-------|
| (8) | *RaCep- i-tú Sa | ‘7’ | > | *pitu |
| | *RaCep- a-telú | ‘8’ | > | *walu |
| | *RaCep-i- Sepát | ‘9’ | > | *Siwa |

The putative PAn forms in (8) are reflected in just one Austronesian language, Pazih (*xasəbidusa* ‘7’, *xasəbituru* ‘8’, *xasəbisupat* ‘9’), and are additive compounds (‘5 + 1’ etc) of which the first constituent is *xasəp* ‘5’, reflecting *RaCep. The latter is also reflected in Saisiyat [*asəb*],³⁴ Favorlang *achab*, Taokas *hasap*, Hoanya *hasip*, but these languages do not share the additive forms found in Pazih.

The comparative method assumes that sound changes are regular, but each of the three derivations in (8) implies a different set of changes. First, *-p- is preserved as *p-

³³ Sagart's subgroups are shown in the left-hand column of Table 6. All languages occurring at or beneath a subgroup label are said to belong to that subgroup, i.e. subgroups are nested.

³⁴ The forms I use here are from modern descriptions instead of the older sources used in some cases by Sagart (2004).

in ‘7’, but has become *w in ‘8’ and ‘9’.³⁵ Second, to account for loss of *-te- in ‘8’, stress is assigned to the forms on the left (PAN stress is otherwise not reconstructable; Blust 1997) and unstressed-syllable deletion is assumed, yielding *walu from *...patelú. If the changes that affect ‘8’ also affected ‘9’, they would yield †*wiwat from *...piSepát in ‘9’, but the outcome instead is *Siwa. Associated with this is the fact that in the forms for ‘7’ and ‘8’ pruning from the left deleted *RaCe... but in ‘9’ *RaCepi.... Third, *Siwa also entails final-consonant deletion, whilst ‘7’ reflects final -CV deletion, neither of which is otherwise attested in Formosan languages.

There are also problems with the evidential basis of Sagart’s reconstructions. He replaces conventionally reconstructed *duSa ‘2’ with *tuSa on the basis of reflexes in Amis, Puyuma and Thao. However, following Tsuchida (1976:153) Ross (1992, 2011) argues that the reconstruction is *DuSa, of which Puyuma *qua* is and Thao *tufa* arguably is the regular reflex.³⁶ Thus only the Amis form reflects a possible *tuSa. Sagart also argues that the final *-p of *RaCep was voiced intervocalically in Pazih, but it is more likely that the earlier form was *RaCeb, with *-b, attested by the Pazih compounds and by Favorlang *achab* and Saisiyat *Lasəb*, and that the final -p of the Pazih, Taokas and Hoanya forms for ‘5’ reflects crosslinguistically common final consonant devoicing. Thus the forms for ‘7’ and ‘8’ which follow from this discussion are *RaCeb-i-DuSa and *RaCeb-a-telu. By the criteria usually required by practitioners of the comparative method, the evidence supporting the PAN reconstructions in (8) is thus meagre in the extreme.

Sagart (2004:416, 419) offers three data fragments in support of left-pruning. These are difficult to evaluate, as they are from manuscripts which tell us nothing about the phonetic values of the transcriptions or the contexts from which they are drawn.

The first is a set of forms from ‘a now-extinct variety of Pazih’:³⁷ *boudah* ‘6’, *bidousut* ‘7’, *bitouro* ‘8’ and *bissoupat* ‘9’. Sagart recognizes ‘6’, ‘8’ and ‘9’ as pruned cognates of Pazih *xasebuza* ‘6’, *xasəbituru* ‘8’, *xasəbidusa* ‘9’, as recorded by Li & Tsuchida 2001. However, left-pruning here is regular, and offers no support for the idiosyncratic derivations in (8).

The second fragment is Luilang *patulu-nai* ‘8’, the first part of which Sagart derives from pruned *...patelú. However, this leaves *satulu-nai* ‘9’ unaccounted for, and an at

³⁵ Sagart (2004:419) suggests that Amis *falú* ‘eight’ supports his claim that *w- in *walu ‘8’ reflects earlier *balu (a misprint for *palu?), but the data in Tsuchida (1981) show that the Proto Amis form was *waLu and that *falú* is a recent and irregular development.

³⁶ Cf. Thao *qutaʃ* < PAN *quDaS ‘grey hair’.

³⁷ Sagart cites these forms from Imbault-Huart (1893:319), who copied them from Thomson (1877), itself apparently a translation from English. Sagart instead cites Thomson (1873) but I cannot find these forms in the latter.

least equally reasonable inference is that *patulu-nai* and *satulu-nai* are subtractive forms ‘minus 2’ and ‘minus 1’, where Luilang *sa-* is ‘one’ (Ferrell 1969:407).

The third fragment is Makatao Siraya *sipat* ‘9’, which Ogawa apparently collected from a non-native policeman (Tsuchida & Yamada 1991:38). It looks like a left-pruned reflex of *RaCep-i-Sepát, but other sources listed by Tsuchida and Yamada have either a reflex of *Siwa or the non-cognate *matuda*. The source appears very unreliable, giving *tasa* for ‘3’ (where other sources have a reflex of *telu) and *supa* for ‘4’ (lacking the *-t* of *Sepat that is allegedly retained in *sipat* ‘9’).

In comparison with the level of regularity evident in sound change across the lexicons of Formosan languages, the changes implicit in (8) are extremely improbable. A reading of Sagart’s (2004:416–421) arguments in support of (8) reveals more arbitrarily assumed changes in addition to the three listed above. Sagart seeks to justify the arbitrariness of these sound changes on the basis of a ‘drive to disyllabism’ which would have applied to very few forms. Such a drive is indeed manifest in Formosan and other Austronesian languages, but not with the ferocity posited in (8) and not with application to compounds (Blust 2007). What we do find, rather infrequently, is deletion of the vowel of an unstressed syllable between a nasal and a stop, converting a trisyllable into a disyllable, as in reflexes of PAn *paŋudal ‘pandanus’:

- (9) a. **Saa** *paŋtal* **Sai** *paŋran* **PAta** *paŋdan **Kav** *paŋzan* **PMP** *paŋdan
 b. **Puy** *paŋuɖal* **PRuk** *paŋuɖale **Pai** *paŋuɖal* **Sir** *paŋadal* **Ami** *paŋizal*

Deletion has not affected the reflexes in (9b), i.e. it occurred at a later stage than that posited for (8).

What is at stake here is not whether the innovations in (8) satisfy the ‘no copying’ and ‘no independent innovation’ conditions, but whether the innovations occurred at all. By the usual assumptions of the comparative method, which are based on observation of what is usual in language change, it is very unlikely that the innovations in (8) did occur, and this undermines hypothesis A. Instead, the resemblances, such as they are, between the forms on the left in (8) and the corresponding forms on the right are probably due to chance (Winter 2010:283–284).

Sagart’s posited tree (Figure 4) requires that the numerals were innovated in the order *pitu ‘7’ > *lima ‘5’ > *enem ‘6’ > *walu ‘8’/*Siwa ‘9’. As Winter (2010) mentions, this would give rise to typologically odd counting systems along the way. For example, Proto Pituish would have *pitu ‘7’ but retain additive numerals on either side, a system which is apparently non-existent among the world’s languages (Eugene Chan, p.c.).³⁸

³⁸ For supporting data, see <http://lingweb.eva.mpg.de/numeral/>.

The innovation which defines the lowest-order grouping in Figure 4, Muish, did not occur at the node to which Sagart assigns it. Muish is so named because its members are said to reflect that part of Blust's (1977) 'second politeness shift' whereby the 2PL genitive enclitic PAn **=mu* came to be used as 2SG genitive. Blust (1977) claims this as a defining innovation of Malayo-Polynesian, but Sagart promotes it to a defining innovation of a subgroup that also includes Northeast Formosan and Tai-Kadai. I will not consider the latter here. Of immediate concern is the claim that the Northeast Formosan languages Ketagalan and Kavalan are Muish. The basis of Sagart's claim is a sentence in Li (1995:667):

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

The text to which Li refers does indeed contain *-imu* as a 2SG possessor pronoun, but after examining further texts Li (1996:179, 1999:639, 643, 663) concludes that in the Basay variety of Ketagalan *[i]mu* was the 2PL form and *[i]su* the 2SG. That is, the politeness shift had not occurred in Basay. It has also not occurred in Kavalan, and it therefore did not occur in Proto NE Formosan. What Li observed in one Ketagalan text does not reflect the politeness shift in a putative Proto-Muish but rather, perhaps, an occasional polite use of a plural for a singular. The 'second politeness shift' reconstructable for PMP was in fact a rather complex set of changes (Ross 2006), and a Muish subgroup would need to reflect the set, rather than a single plural-to-singular shift. There is thus no evidence for the Muish node in Figure 4. Instead, NE Formosan and PMP would simply join the list to the right of the Walu-Siwaish node.³⁹

4.3.2 Hypothesis B

For Sagart the innovations in (8) establish directionality, i.e. they point to hypothesis A rather than B. If these innovations did not occur, then we should consider hypothesis B, that **puluq* etc are PAn forms that have been lost in certain Formosan languages. Hypothesis B inverts the relationship between the two parts of Table 6: it is the forms in white slots that are innovatory.

³⁹ Sagart's Muish also includes the ancestor of the Tai-Kadai languages. If these are taken into account, then the Muish node survives, but is the mother of only the ancestor of the Tai-Kadai and of PMP.

Hypothesis B envisages the replacement of unit forms for 6-9 by additive compounds. From what we know about similar developments in other Austronesian languages, such events are quite common. Sagart mentions languages in the northern Philippines, northern Vanuatu and New Caledonia in this regard. A quick survey of Lynch, Ross & Crowley (2002) shows that of the 43 Oceanic languages of which sketches are provided in the volume, 21 have compound or compound-derived forms for 6-9.⁴⁰ The distribution of these languages is patchy. They do not occur in the Micronesian or Central Pacific subgroups of Oceanic, but they occur everywhere else, interspersed with languages that retain reflexes of PAn *enem ‘6’, *pitu ‘7’, *walu ‘8’ and *Siwa ‘9’. Thus one can claim that a hypothesis like B is reasonable. But we need to ask how the replacement of unit numerals by compounds happens. The answer appears to be that it is the result of contact (Blust 2005:552-556),⁴¹ in the Formosan case perhaps with speakers of the language(s) of the pre-Austronesian Changpin Culture.

Table 7 illustrates the discussion below. Forms from modern descriptions are used where possible, most as cited by Li (2006), as their phonetics and phonologies are known. Adelaar (2011) is used for Siraya. Retentions from PAn are shown in shades of grey. Table 7 shows a reduction of the numeral paradigm such that the monomorphemic forms for ‘6’ to ‘9’ in the grey cells are largely replaced by compounds in the unshaded cells.

⁴⁰ The 21 languages are, arranged by subgroup, Sobei (Sarmi), Arop-Lokep, Kaulong, Yabem, Takia, Kairiru (North New Guinea), Gapapaiwa, ‘Ala’ala (Papuan Tip), Bali-Vitu, Banoni (Meso-Melanesian), Mwotlap (N Vanuatu), SE Ambrym, Lamen, Sakao, Vinmavis, Port Sandwich (C Vanuatu), Anejom, Sye (S Vanuatu), Cèmuhî, Xârâcùù (New Caledonia), Iai (Loyalties).

⁴¹ Blust (2005) argues that the presence of languages with compounds for 6-9 in SE Melanesia (Vanuatu and New Caledonia) shows that Papuan languages were already spoken there when Oceanic speakers arrived. The archaeology, however, indicates that Oceanic speakers were the first arrivals in SE Melanesia. Pawley (2006:247-248) suggests that these first arrivals were accompanied by Papuan speakers from the New Guinea region and that contact effects thus continued after the settlement of SE Melanesia.

Table 7: Innovations in the numerals 5-10 in Formosan languages
(see text for explanation)

PAn	*lima ‘5’	*enem ‘6’	*pitu ‘7’	*walu ‘8’	*Siwa ‘9’	*puluq ‘cardinal 10’	*ma[ŋ]saL ‘-teen, -ty’ etc
Luilang	[na]lup	[na]tsulup	in-nai	patulu-nai	satulu-nai	isit	...
Pazih	xasəp	xasəbuza	xasəbidusa	xasəbituru	xasəbisupat	isit	isid-u
Saisiyat	rasəb	ʃaiboʃir	ʃaiboʃir o ʔəhæʔ	ka-fpat	ræ-ʔhæʔ		lanpəz
Favorlang	achab	nataap	naito	maaspat	tannacho	zchiətt	...
Taokas	hasap	tahap	yweto	mahalpat	tanasu	ta-isid	...
Hoanya	lima, hasip	<i>mi-nun</i>	pito	<i>mi-alu</i>	<i>a-sia</i>	miata-isi	...
Papora	nema	<i>ne-nom</i>	pitu	<i>mahal</i>	<i>me-siya</i>	me-tsi	...
Thao	rima	ka-turu	pitu	ka-fpat	ta-na-θu		makθin
Mayrinax Atayal	timaʔ	ma-tuuʔ	pituʔ	ma-spat	ma-qisuʔ	puy	maxal-puy
Paran Seediq	rima	m-teru	m-pitu	m-sepac	m-ŋariʔ		maxal
Siraya	rīma	ənəm	pītu	kuixpa	matuda	kītiän	äb
Kavalan	rima	ʔnəm	pitu	waru	siwa	stRay	Rabtin
Basay	tsjima	anəm	pitu	wasu	siwa	labatan	...
Duhtu Tsou	rimo	nomə	pitu	voru	sio		maskə
Saaroa	ku-lima	k-ənəmə	ku-pitu	ku-alu	ku-sia	ku-maatə	laiʔa
Kanavu	liʔma	nəʔmə	piʔtu	áalu	síiya		máanə
Bunun	himaʔ	num	pituʔ	vauʔ	sivaʔ		macʔan
PRukai	*Lima	*ənəmə	*pitu	*vaLu	*banjatə	*puLuku	*maŋə-salə
Puyuma	ʃima	nəm	pitu	waʃu	iwa	puʃuʔ	mukʔəp
Amis	lima	ʔənəm	pitu	falu	siwa	puluq	muʔtəp
Paiwan	ʃima	unəm	pitʃu	aʃu	siva		puʃuʔ
PMP	*lima	*enem	*pitu	*walu	*siwa		*puluq

Based on Li’s (2006) findings, there are two columns for ‘10’ in Table 7. PAn *puluq was a cardinal, and PAn *ma[ŋ]saL was used for counting objects and in ‘11’ to ‘19’. Reflexes of *ma[ŋ]saL are boxed and shown in a shade of grey to indicate that they are not innovative. Proto Rukai reflected *maŋ(e)-sa-L, other languages *ma-sa-L. As the data in Li (2006) indicate, PAn decade terms were formed with *ma-...-[e]L,⁴² e.g. *ma-puSa-L ‘20’, *ma-telu-L ‘30’ (< *telu ‘3’), *ma-Sepāt-eL ‘40’ (*Sepāt ‘4’), and so on, and *ma-sa-L ‘10’ (< *sa ‘one’) belongs to this series (Tsuchida 1976:205).

Once the two terms for ‘10’ are explained, Table 7 shows that numeral systems in Formosan languages form two areas. In one, from Kavalan to the bottom of the table,

⁴² This construction is reflected in Atayal, Seediq, Tsou, Kanakanavu, Saaroa, Rukai, Bunun, Puyuma, Thao and Saisiyat (Li 2006). The replacement of *duSa ‘two’ by *-puSa- here remains unexplained.

PAn numerals are largely preserved in a geographic area that covers the east coast and mountain groups with the exception of Atayalic (Atayal and Seediq). In the other, from Seediq to the top of the table, numerals from ‘6’ to ‘9’ tend to be replaced by compounds in an innovative area covering the northwestern plain and the mountain valleys facing the plain.

In the innovative area replacement follows a partially discernible pattern. The even numbers ‘6’ and ‘8’ are, as Li (2006) points out, replaced by ‘double 3’ and ‘double 4’, as shown in (10) and (11), where ‘double’ is encoded by the same pair of affixes as marks many stative verbs in Formosan languages, namely *ma- ‘actor voice’ and *ka- ‘undergoer voice’ (Huang 2000, Zeitoun 2000, Zeitoun & Huang 2000).⁴³ The odd number ‘9’ is expressed subtractively by PAn *sa-na-Cu, from *sa ‘one’ + *na GENITIVE DETERMINER + *Cu ‘that’ (Li 2006:141), shown in (12). Reflexes of the forms in (10), (11) and (12) are shown in bold in Table 7. Their extent is interrupted by the italicized Hoanya and Papora forms, which Sagart attributes to copying from Kanakanavu or Saaroa.

- (10) a. PAn *ma-telu ‘6’ (*telu ‘3’) > **Ata** *ma-tuu?* **Paran Seediq** *m-teru*
 b. PAn *ka-telu ‘6’ > **Tha** *ka-turu*
- (11) a. PAn *ma-Sepat ‘8’ (*Sepat ‘4’) > **Ata** *ma-spat* **Paran Seediq** *m-sepac*
Fav *maa-spat*
 b. PAn *ka-Sepat ‘8’ > **Tha** *ka-fpat* **Sai** *ka-fpat*
- (12) PAn *sa-na-Cu ‘9’ > **Tha** *ta-na-θu* **Tao** *ta-na-su* **Fav** *tannacho*

It is probable that Saisiyat *ræ-ʔhæ?* ‘9’ (*ʔæhæ?* ‘one’) is also subtractive, and perhaps also Luilang *satulu-nai* (§4.3.1).

This account goes some way towards explaining why *pitu ‘7’ has been retained in the Atayalic and Western Plains languages when ‘6’, ‘8’ and ‘9’ have been replaced. There is a straightforward subtractive strategy to encode the odd number ‘9’ with a compound (10-1), but not ‘7’. Saisiyat resorts to the extraordinary *faibofir o ʔæhæ?* ‘6 + 1’ to encode it, whilst the Central Plains and Atayalic languages retain *pitu.

The distributions of innovations in the innovative area border on chaos, and one suspects that the italicized items are not the only copies here. The lexical replacements evident in the reflexes of apparent *RaCeb ‘5’ and *iCi(d,Z) ‘10’, boxed with broken lines in Table 7,⁴⁴ are coterminous neither with each other nor with the innovations above.

⁴³ The connection between stativity and ‘double’ is not clear to me, but the forms in (10) and (11) must have originated as verbs.

⁴⁴ Luilang *na-lup* ‘5’ may also reflect *Raceb, but there is not enough evidence to reconstruct Luilang historical phonology.

This is not surprising, however. For a considerable period of time after the dispersal of PAn, its daughters must have formed a network in which discrete subgroups were the exception rather than the rule (§2.1; also Li 2006:139) and the spread of innovations would not have been prevented by subgroup boundaries. Also unsurprising is the fact that the area of overlapping isoglosses is centred on the northwest plain, where links in the network are likely to have lasted longer than they did in the mountains.

The innovative northwestern area falls in turn into two sub-areas. The first contains the Atayalic and Western Plains languages, where *pitu ‘7’ is retained. The second consists of three disparate systems: Luilang (which defies analysis), Saisiyat (which is odd), and the uniquely additive system of Pazih.

On the face of it, Siraya forms a typological transition between the two areas in that it retains *enem ‘6’ as well as *pitu ‘7’, but it is difficult to interpret the Siraya forms for ‘8’, ‘9’ and ‘10’ without data from (now lost) surrounding languages. Geographically Siraya, in the southwest, belongs to neither of the two areal groups and quite possibly has a history somewhat different from either.

The conflict between the NAn hypothesis and Sagart’s subgrouping is more far-reaching than are those between NAn and either Blust’s or Ho’s subgroups (§4.4). Despite their enormous morphosyntactic differences, Puyuma, Tsou and Rukai, each a primary branch under the NAn hypothesis, all fall under Sagart’s third-order Walu-Siwaish, whilst NAn is split across all his first-, second- and third-order subgroups. It is for this reason that I have provided an alternative account of the facts Sagart presents.

4.4 Pulling threads together

The examination in §4.1 and §4.2 of Blust’s and Ho’s subgroupings suggests that there is evidence for the Western Plains, Atayalic and Northeast Formosan subgroups. These fall within NAn (§3), together with Paiwan, Bunun, Pazih, Saisiyat, Siraya and Amis, for which no obvious lower-order subgroupings emerge on the basis of shared innovations. Table 5 shows a number of innovations that criss-cross the languages and subgroups within NAn in a way that suggests that PNAn first diversified into a linkage rather than splitting directly into discrete languages. The reanalysis of Sagart’s data in §4.3.2 suggests that the linkage lasted longer on the northwestern plain and the mountain valleys facing the plain than in the southern and east coast region, which in any case contains languages of all four primary subgroups recognized under the NAn hypothesis.

None of this is surprising. In view of the expectation that PAn would have diversified into a linkage (§2.1), it is the existence of NAn, a large primary subgroup, that is somewhat surprising. However, the boundaries between the four primary branches under the NAn hypothesis all lie within the southern part of the cordillera, and one may reasonably

infer that the initial split into four very different languages occurred because of the settlement of isolated mountain valleys, whilst the diversification of PNA into a linkage happened when its speakers occupied the plains to the west of the cordillera.⁴⁵

The single point of conflict between Blust's and Ho's subgroupings and the NAn hypothesis is the Tsouic subgroup. Under the NAn hypothesis the Tsouic subgroup must be broken up into Tsou, a primary branch of Austronesian in its own right, and Kanakanavu-Saaroa. As the Tsouic subgroup has been almost universally accepted by scholars, including Blust, Ho and Sagart, §5.1 is devoted to showing how different Tsou is from Kanakanavu and Saaroa, and §5.2 to an examination and rejection of the evidence offered for Tsouic. In §5.3 I briefly discuss Rukai-Tsouic, as Ho and Sagart include it in their subgroupings.

5. Tsou, Tsouic and Rukai-Tsouic

A major issue for the NAn hypothesis is the possible existence of a Tsouic subgroup. Because Tsou is not a NAn language but Kanakanavu and Saaroa both are (and are apparently closely related), a Tsouic subgroup is incompatible with the NAn subgroup. Tsouic has been accepted without further argumentation by most scholars since the appearance of Tsuchida's groundbreaking 1976 work on Tsou, Kanakanavu and Saaroa. Blust (1999:52) writes about the Tsouic subgroup:

The fundamental evidence for a Tsouic subgroup has been presented by Tsuchida (1976). Although some writers have questioned the unity of Tsouic (Harvey 1982:90), I accept it as established on the basis of Tsuchida's extensive documentation.

Ho (1998) bases his acceptance of the Tsouic subgroup on the mergers of PAn *k with *g and of PAn *D with *z.

The Tsouic subgroup has had few detractors. Blust mentions Harvey (1982:89-90), who adduces just two pieces of phonological evidence for Tsouic, the merger of PAn *C and *d and the reduction of word-final *-an to -a, when final consonants are otherwise retained. Harvey finds this unconvincing evidence of a Tsouic subgroup, and I will show below (§5.2.4, §5.2.5.1) that neither of these innovations is a shared innovation of

⁴⁵ This places the PAn homeland in the southern mountains. It is important to note, though, that this does not mean that the southern mountains were the area first occupied by arriving pre-PAn speakers. It is simply the oldest homeland that can be reconstructed on the basis of data from the twenty or so languages for which data survive.

Tsou, Kanakanavu and Saaroa. The similarities that have been noted among the Tsouic languages and between them and the ‘Lower Three Villages’ Rukai dialects, Harvey suggests, may be due to contact. He alludes to Ferrell (1969:36-29), who notes the cultural traits that closely bind speakers of Tsou, Kanakanavu and Saaroa together and separate them from speakers of other Formosan languages. Regarding the languages Ferrell notes:

Grammatically, lexically and phonologically, Tsou is by far the most aberrant of all Formosan languages, leading us to suspect that its separation from the ancestors of the other Formosan languages was at a very remote period indeed. However, as will be discussed later, Tsou linguistic peculiarities are shared only to a limited extent by Kanakanabu and Saaroa. (Ferrell 1969:39)

The later discussion is worth quoting in full:⁴⁶

Kanakanavu and Saaroa appear to have few of the structural complexities found in Tsou, and their lexical similarities to neighboring Paiwanic languages make them appear to form a possible link between Tsou and the Paiwanic languages. Kanakanavu and Saaroa are very small ethnic groups, with about 160 and 278 speakers respectively at present, who have had intense contact and admixture with neighboring groups not only in the present century but at least as early as the 17th century. The “Tevorang” region (...) near the homeland of the Saaroa, has been a melting pot particularly for Saaroa, Rukai and Siraya for well over 300 years that we know about. The result has been that although the Saaroa are culturally Tsouic, their vocabulary resemblances to Siraya and Rukai are so numerous that one may wonder whether Saaroa is indeed a Tsouic language, with extensive influences from neighboring Paiwanic languages, or whether it may in fact be a Paiwanic language with heavy Tsouic overlay. Superficial phonological examination leads me to continue the traditional classification of Saaroa and Kanakanavu as Tsouic, and to consider Tsouic a discrete grouping. It is obvious that these questions ... cannot be decided until structural and phonological studies in depth are completed. (Ferrell 1969:67-68)

⁴⁶ Ferrell refers here to ‘Paiwanic languages’. For him Paiwanic includes all Formosan languages other than Tsouic and Atayalic.

I show in §5.2 that Ferrell's assessment was remarkably insightful. The evidence adduced for a Tsouic subgroup does not stand up to close inspection but does suggest intense and longstanding contact between Tsou, Kanakanavu and Saaroa.

5.1 Tsou is different

Scholars have only rarely followed Ferrell in drawing attention to how different Tsou is from Kanakanavu and Saaroa, and indeed from all other Formosan (and, one might add, Austronesian) languages. Exceptions are Starosta (1985) and H. Y. Chang (2006). Starosta noted that Tsou lacks second-generation verbal forms (it also lacks nominal reflexes of second-generation affixes other than *-an) and was the first to suggest that Tsou is a first-order branch of Austronesian. Chang provides a list of differences between Tsou and Kanakanavu-Saaroa which is the backbone of the present section. Where historical linguists have recognized that Tsou is different, they have tended to assume that the differences are due to innovations in Tsou that have occurred since the break-up of presumed Proto Tsouic. I am not aware that this view has appeared in print, but H. Y. Chang (2006:580) cites personal communications to this effect, and I subscribed to this view myself before taking a closer look at Tsou, Kanakanavu and Saaroa in their diachronic context.

Here I will briefly list the differences between Tsou on one hand and Kanakanavu, Saaroa and other Formosan languages on the other, then suggest that there is enough circumstantial evidence to infer that Ferrell was right to suspect of Tsou 'that its separation from the ancestors of the other Formosan languages was at a very remote period indeed.'

5.1.1 Consonant clusters and allomorphy

A major phonological difference between Tsou and most other Formosan languages is that Tsou allows more consonant clusters than most other Formosan languages (Wright 1996, Zeitoun 2005:261-262).⁴⁷ These are the result of vowel losses described by Tsuchida (1976) and summarized in §5.2.3. These changes have also led to considerable verb-stem allomorphy (Tsuchida 1976:102-110). The only other Formosan language to display similar phenomena is Atayal (Egerod 1965).

⁴⁷ A referee points out that Thao has a substantial inventory of consonant clusters, and Elizabeth Zeitoun (p.c.) notes that this is also true of Maga Rukai.

5.1.2 The morphology and syntax of voice

H. Y. Chang (2006) describes six morphosyntactic differences between Tsou and other Formosan languages.⁴⁸ The first of these (pp.566-570) is the difference between verbal voice affixes in Tsou and other Formosan languages. The fifth (pp.576-578) is the fact that all Tsou independent and many dependent clauses begin with a preverb (auxiliary). Taken together, they represent the most pervasive morphosyntactic difference between Tsou and other Formosan languages: in Tsou all independent and many dependent clauses begin with a preverb, followed by a verb reflecting one of the PAN dependent forms.

- (13) a. *mo=∅* *m-osi* *ta* *paŋka* *to* *emi* *?o* *amo*
 AV.RLS=3S.NOM AV-put OBL table OBL wine NOM father
 ‘Father put some wine on the table.’
- b. *i=si* *si-a* *ta* *paŋka* *to* *amo* *?e* *emi*
 UV.RLS=3S.GEN put-UIP OBL table OBL father NOM wine
 ‘Father put the wine on the table.’
- c. *i=si* *si-i* *to* *emi* *to* *amo* *?e* *paŋka*
 UV.RLS=3S.GEN put-UVL OBL wine OBL father NOM table
 ‘Father put the wine on the table.’ (Zeitoun 2005:266)

As the forms shown in Table 8 indicate, the dependent verb form encodes voice. Some preverbs also encode the lesser distinction between actor and undergoer voice, but not the three subtypes of undergoer voice, as (13b) and (13c) show (Zeitoun 2005:265-272).

Table 8: Tsou verbal morphology, illustrated with forms of Tsou *(o)si* ‘put’

		UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Dependent	<i>M-stem</i>	STEM- <i>a</i>	STEM- <i>i</i>	STEM- <i>[n]eni</i>
	<i>m-osi</i>	<i>si-a</i>	<i>si-ia</i>	<i>si-eni</i>

⁴⁸ I omit the sixth of these, the emergence of 3SG bound pronouns in Tsou (H. Y. Chang 2006: 578-579), because (i) I am uncertain whether PAN lacked a 3SG enclitic pronoun; and (ii) the tendency for third person pronominals to emerge out of demonstratives is so widespread that what has happened in Tsou is unexceptional.

A comparison of Table 8 with Table 1 shows that Tsou retains only PAn dependent forms. It has lost all independent forms together with the constructions in which they occurred. It is thus unlike other Formosan languages, of which

- A. Puyuma comes close to retaining the reconstructed PAn system, but has lost all preverbs except the negator;
- B. Rukai has lost all undergoer voice forms and has acquired a new passive (Zeitoun & Teng 2009);
- C. all others, including Kanakanavu and Saaroa (H. Y. Chang 2006:568-570), have undergone the Nom-into-V innovation.

Since I claim that Puyuma is the only Austronesian language that closely reflects the PAn system of verbal morphology, one might instead argue that Proto Austronesian split initially into Tsou and X, an ancestor of all other Austronesian languages, and that the independent verb forms which are not reflected in Tsou were innovated in X. Puyuma would then reflect X rather than PAn. Whilst this order of splits is on other grounds quite plausible, the paradigmatic patterning noted in §3.1 is evidence that at least unreduplicated Puyuma independent forms were already present in PAn. If the affixes **M-...-a*, **-aw*, **-ay* (optative/hortatives in Table 2) were derived from a suffix **-a* plus **M-*, **-u* and **-i* respectively (imperatives and dependents), then we need to accept that **-a*, **-aw*, **-ay* were contemporaneous with zero, **-u* and **-i*, and that Tsou has lost them. The alternative requires the unlikely claim that X inserted **-a* before the existing imperative and dependent suffixes.

5.1.3 Nominalizing morphology

The second feature noted by H. Y. Chang (2006:570-573) is the absence from both Tsou verbs and nominalizations of the PAn perfective infix **<in>* and its prefixal allomorph **ni-*. Its absence from verbs is not surprising, as PAn **<in>* formed perfective nominalizations, and Tsou has never undergone the Nom-into-V innovation that turned nominalizing morphemes into second-generation verbal affixes in NAn languages (§3.1; Ross 2009). But as Chang points out, Tsou is the only Formosan language in which PAn **<in>* is not reflected in nominalizations. In fact, Tsou does not appear to have the distinction between perfective and imperfective nominalizations that occurs in other Formosan languages.

Also absent from Tsou nominalizations are reflexes of the PAn patient suffix **-en* and the circumstance prefixes **Si-* and **Sa-*. Instead we find the following nominalizing morphology (Szakos 1994:73-75), some of it derived from known PAn morphemes as indicated in (14), some of it not.

(14)	Tsou	PAn
actor	<i>to</i> -STEM	* <i>ta</i> -STEM (7)
person, location	<i>le</i> -STEM	?
patient	STEM- <i>a</i>	*STEM-an ⁴⁹
location, realis	STEM- <i>a</i>	*STEM-an (Table 1)
location, irrealis	<i>Co</i> -STEM- <i>a</i>	*Ca-STEM-an (Table 1)
location	<i>to</i> -STEM- <i>a</i>	* <i>ta</i> -STEM-an (5), (7)
circumstance, location	<i>o</i> -STEM- <i>a</i>	* <i>a</i> -STEM-an ?
state	STEM, <i>hʔo</i> -STEM	?

Starosta (1985) remarks that if the ‘absent’ forms had ever occurred in Tsou, we would expect them to be reflected at least in lexicalized nominals, and yet they are apparently not found. This gives us a historical conundrum which is discussed in §5.1.6.

5.1.4 Serial verb constructions

Tsou serial verb constructions exhibit what H. Y. Chang (2006:573-574) calls the ‘focus harmony restriction’, i.e. all verbs in a single serial verb construction must have the same voice:

- (15) a. **mi**=ʔo kaeb-**u** bon-**u** ta tacumʌ
 AV=1s like-AV eat-AV OBL banana
 ‘I like to eat bananas.’
- b. **os**=ʔo kaeb-**a** an-**a** ʔo tacumʌ
 UV=1s like-UV eat-UV NOM banana
 ‘I like to eat the bananas.’ (H. Y. Chang 2006:573)

In this regard Tsou is unlike all other Formosan languages except Rukai, which obey a restriction whereby every verb after the first in a serial verb construction must have actor voice, regardless of the voice of the initial verb, illustrated in these Kanakanavu examples:⁵⁰

⁴⁹ Tsou patient nominalizations like *hopeiʔa* ‘prepared food’, *tənəʔva* ‘present, gift of food’, *masasajeava* ‘drink’ (Szakos n.d.) reflect *STEM-an. Rukai patient nominalizations also do so, but Katipul Puyuma patient nominalizations reflect *STEM-en (§3.3), casting doubt on the PAn form, which is left blank in Table 1.

⁵⁰ Mantauran Rukai is unlike either Tsou or other Formosan languages. It has only an active/passive distinction. Serialized verbs are usually both active, but Zeitoun (2007:404) also provides examples of active + passive and passive + passive.

- (16) a. urupaca=ku aratin k<um>a-kun
 use:AV=NOM.1S chopstick <AV>IMPF-eat
 ‘I use chopsticks to eat.’
 b. urupac-**un**=maku aratin isi k<um>a-kun
 use-UVP=GEN.1S chopsticks this <AV>IMPF-eat
 ‘I use this chopstick to eat.’ (Wu 2006:125)

5.1.5 The morphological causative

The ‘normal’ Formosan causative is illustrated in (17) with examples from Paiwan. The verb in (17a) is the actor voice of *alup* ‘hunt’. In (17b) the actor voice of the causative verb *pa-alup* is formed by prefixing *pa-* ‘causative’ to the stem *alup* (the AV of a causative verb is not marked by a voice affix). In (17c) the circumstance undergoer voice of the causative verb *pa-alup* is marked as usual with *si-*.

- (17) a. na=alup ti palang ta vavuy.
 PRF=<AV>hunt NOM:PERS:SG Palang OBL wild.pig
 ‘Palang hunted wild pigs.’
 b. na=pa-alup=aken tay palang ta vavuy.
 PRF=AV.CAUS-hunt=NOM:1S OBL:PERS:SG Palang OBL wild.pig
 ‘I made Palang hunt wild pigs.’
 c. ku=s<in>i-pa-alup tay palang a icu
 GEN:1S=<PRF>UVC-CAUS-hunt OBL:PERS:SG Palang NOM this
 a vavuy.
 LINKER wild.pig
 ‘I made Palang hunt this wild pig.’ (A. H. Chang 2006)

The Tsou reflex of the morphological causative, illustrated in (18), differs from the causative in Paiwan and other Formosan languages in three respects. First, the PAn causative prefix **pa-* is reflected irregularly as Tsou *poa-* (with allomorphs *pa-* and *p-*). Second, a Tsou causative verb is almost always in undergoer voice (patient or circumstance; H. Y. Chang 2006:574-575). This means that its subject is either the causee or the patient, but not the causer (Tung 1964:191-192, Starosta 1974:351-362, H. Y. Chang 2006:574-575, C.-W. H. Chang 2008).⁵¹ Third, unlike Paiwan and other Formosan languages, where the causative verb is formed from *pa-* + stem, Tsou *poa-* is attached to a voice-marked form of the verb. Thus the causative verb *poa-bonu* ‘cause to eat’ in (18b)

⁵¹ There are a few exceptions to the subject rule, as certain lexicalized forms with *poa-* may occur in actor voice with a causer subject (S. Huang & H.-J. Huang 2005).

(which occurs in the patient-undergoer voice form *poa-bonɯ-a*, where *-a* is optional) has as its stem the actor voice verb form *bonɯ* ‘eat’ in (18a), and the case-marked arguments in (18a) are carried over into (18b). Similarly in (18d), the causative verb *poa-an-a* ‘cause to be eaten’ (which occurs in the circumstance-undergoer voice form *poa-an-a-neni*) has as its stem the patient-undergoer voice verb form *an-a* ‘be eaten’ in (18c), and the case-marked arguments in (18c) are carried over into (18d).

- (18) a. mo-∅ bonɯ to aveʔu ʔo fokɲe
 AV.RLS=NOM:3S AV.eat OBL rice NOM frog
 ‘The frog ate rice.’ (constructed example)
- b. os=ʔo poa-bonɯ(-a) to aveʔu ʔo fokɲe
 UV.RLS=GEN:1S CAUS-AV.eat(-UVP) OBL rice NOM frog
 ‘I made the frog eat rice.’ (S. Huang & H.-J. Huang 2005)
- c. i=si an-a to pasuya ʔo fʔue
 UV.RLS=GEN:3S eat-UVP OBL Pasuya NOM yam
 ‘Pasuya ate the yam.’ (constructed example)
- d. i=si poa-an-a-neni to pasuya to voyu ʔo fʔue
 UV.RLS=GEN:3S CAUS-eat-UVP-UVU OBL Pasuya OBL Voyu NOM yam
 ‘Voyu made Pasuya eat the yam.’ (S. Huang & H.-J. Huang 2005)

There are one or two clues that suggest that this development represents an innovation in Tsou relative to the structure reflected in other Formosan languages. S. Huang & H.-J. Huang (2005) note that there are lexicalized causatives like *p'onɯ* ‘feed’ in (19) which occur in actor voice and have a structure corresponding to the putative PAn construction reflected in (17b).

- (19) mi=ʔo pʔonɯ to naveu to avʔu
 AV=NOM:1S AV.feed OBL rice OBL dog
 ‘I fed the dog with rice.’ OR ‘I fed rice to the dog.’ (S. Huang & H.-J. Huang 2005)

The synchronic derivation of (18b) from (18a) and of (18d) from (18c) appears to reflect a diachronic process such that the causative structures came into being later than the simple voice forms, and C.-W. H. Chang (2008) suggests that Tsou *poa-* reflects the (perhaps incomplete) grammaticization of an earlier verb.

Kanakanavu and Saaroa have *pa-*stem verbs which behave like those in Paiwan and other Formosan languages (H. Y. Chang 2006).

5.1.6 Why is Tsou different?

Proponents of the Tsouic subgroup necessarily view the differences between Tsou and other Formosan languages as outcomes of innovations that have taken place since Tsou separated from the shared ancestor of Kanakanavu and Saaroa. This view, however, is problematic. Languages do not undergo morphosyntactic change willy-nilly, and there is an increasing body of evidence that radical morphosyntactic changes, i.e. morphosyntactic changes that are not due to incremental grammaticization, are due to contact. Thus Dunn et al. (2011) show that in several major language families there is considerable typological homogeneity among the languages of the family. For Austronesian they show that languages have verb-object and preposition-NP orders across the entire family, except in that part of the family that is located in New Guinea and has been subject to known contact with Papuan languages. There, object-verb and NP-postposition orders prevail. The features in which Tsou differs from other Formosan languages are, with the possible exception of its nominalizing morphology, of a radical nature and the differences between Tsou and other Formosan languages can reasonably be attributed to contact, either between Tsou and a non-Formosan language or between an ancestor of the other Formosan languages and one or more non-Formosan languages. The only non-Formosan languages with which early Formosan languages have come into contact are the languages of the putative pre-Austronesian inhabitants of Taiwan, and we may reasonably assume (§2.1) that these disappeared fairly soon after Austronesian speakers first arrived in Taiwan. This places the divergence between (pre-)Tsou and the ancestors of the other Formosan languages very early in the history of Austronesian and is strong circumstantial evidence against a Tsouic subgroup.

The question to which this leads is, of course, who innovated? Was PAn more like Tsou or more like the ancestors of the other Formosan languages? In the case of the morphology and syntax of voice (§5.1.2), I have provided a reason for thinking that Tsou was the innovator. On the other hand Starosta (1985) points out that the nominalizing morphemes that are missing from Tsou (§5.1.3) ought, if they were ever present, to have reflexes in lexicalized nominalizations—but they don't. This implies that the morphemes that are missing from Tsou were innovated (through grammaticization) in a language ancestral to all other Formosan languages after it had separated from Tsou. The differences in serial verb constructions (§5.1.4) are neutral with regard to the competing hypotheses.

There is no contradiction here. It is a reasonable hypothesis that PAn split into pre-Tsou and a language ancestral to Puyuma, Rukai and the NAn languages. At least this is how the split would be viewed under the family tree model, inserting an additional node in Figure 1. But a more plausible scenario is that PAn diversified into an early linkage

(§2.1), and then that the part of that linkage that represented pre-Tsou became isolated in its mountain home, whilst the rest of the linkage remained unbroken. After pre-Tsou became isolated it underwent the restructuring of the morphology and syntax of voice outlined in §5.1.2 and made the changes to the morphological causative described in §5.1.4. The rest of the Austronesian linkage then innovated the nominalizing morphemes *⟨in⟩, *-en and *Si-/Sa (§5.1.3), while pre-Tsou at some point innovated the morphemes listed without a PAn source in (14).

If the hypothesis of the previous paragraph were adopted, it would necessitate an additional node in Figure 1 and a simplification of the nominalizing morphology in Table 2. However, the hypothesis hangs by such a slim thread, depending on nominalizing morphology alone, that it must remain an interesting speculation.⁵²

The circumstantial evidence of the differences between Tsou and other Formosan languages sets the scene for a demonstration in §5.2 that there is no compelling evidence for a Tsouic subgroup.

5.2 A Tsouic subgroup?

5.2.1 Previous work

The major work on Tsouic languages is Tsuchida (1976), but he does not demonstrate that Tsouic is a subgroup. In his introduction he writes that on the basis of lexicostatistics and lexical comparison (mainly Dyen 1963, Ferrell 1969 and Li 1972) certain subgroupings of the Formosan languages—Tsouic, Paiwanic and Atayalic—‘have been generally accepted and seem to be indisputable’ (1976:9-10). Lexicostatistics has been shown to deliver false results (e.g. Blust 2000), and fails to distinguish between innovations and retentions. Distinguishing them in Taiwan is in any case particularly difficult because of the root-node limitation: we often cannot tell with certainty what is inherited and what is innovated. Of Tsuchida’s sources, only Li (1972) employs the comparative method. He mentions two alleged Tsouic innovations, loss of PAn *j and the merger of *d/*z. However, *j is lost in Saaroa and Kananavu only adjacent to *i (Tsuchida 1976:223), and the *d/*z merger occurs so widely in Formosan languages (Table 5) that it does not meet the ‘no independent innovation’ condition.

Of the three subgroups that Tsuchida describes as ‘generally accepted’ Paiwanic is now generally rejected, and I contend that Tsouic should also be rejected. Although Tsuchida takes the Tsouic subgroup as a given, he describes a number of innovations which may be taken as evidence for Tsouic.

⁵² See Ross (2012) for further evidence for this hypothesis.

5.2.2 Sibilant dissimilation

The strongest of these is sibilant dissimilation, a conditioned change whereby PAn *C (presumably [ts]) becomes *t* in the three Tsouic languages in words that also include a sibilant, as shown in (20), but remains unchanged as *c* elsewhere (Tsuchida 1976: 148-151).⁵³

- (20) a. **PAn** *CaliS ‘rope, string’ > **Tso** tresi⁵⁴ **Kan** talísi, **Saa** li-ki-taŋí-a
RukMag tési **RukTon** taísi (cf **RukTan** caŋisi **RukBud** cáŋisi)
 b. **PAn** *CiŋaS ‘food particles between teeth’ > **Tso** ru-ŋtose ‘remove food particles’ **Saa** ʔi-u-tiŋa-a, **RukMag** mu-tiŋasə ‘remove food particles’
RukTon mua-tiŋasə ‘remove food particles’ (cf **RukBud** mua-cíŋasə
RukMan mu-ciŋaʔa)
 c. **PAn** *CaqiS ‘sew’ > **Tso** tɤmɤeʔsi/tʔes-a **Kan** tɤmɤa-taʔisi (**RukTan**
RukBud caísi)
 d. **PAn** *Caŋis ‘weep’ > **Tso** tɤjis-i **Kan** t-um-á-taŋi **Saa** tɤmɤaŋi/t-um-a-taŋii

Three other etyma, shown in (21), display the same change but without the conditioning, i.e. as idiosyncratic lexical changes.

- (21) a. **PAn** *CugCug ‘be bumped on head’ > **Saa** m-utu-tukutuku **RukMag**
mu-tgútgu (**Pai** tsɤmɤug-tsug ‘hit at s.t.’)
 b. **PAn** *LaCeŋ ‘vegetables’ > **Kan** natéŋe **Saa** lateŋe **RukMag** [cǝŋǝ
RukTon lacéŋe **RukBud** láceŋe
 c. **PAn** *Caqi ‘faeces’ > **Tso** tʔee **Kan** táaʔi **Saa** tiiʔi (**RukMag** ckee,
RukTon RukTan RukMan caki, **RukBud** cáki)⁵⁵

In another idiosyncratic case both PAn *tebuS and PAn *CebuS ‘sugarcane’ can be reconstructed (Puyuma, Paiwan, Thao, Taokas, Favorlang-Babuza, Pazih and Saisiyat forms reflect *t-, Papora and Atayalic *C-) and the three Tsouic languages reflect *t.

⁵³ PAn *s was probably [c], *S perhaps [ʃ] or [ʂ] (Ross 2011).

⁵⁴ The phoneme /r/ survives only in the Duhtu dialect of Tsou. In the Tfuea and Tapangu dialects it is replaced by /e/. Hence Duhtu *tresi* corresponds to *teesi* in the other two dialects.

⁵⁵ The PAn term for ‘faeces’ had two variants, one with *-q-, widely reflected in Austronesian languages, and one with *-k-, reflected in Rukai and Bunun *takiʔ*, Pazih *saiik* and in various Philippine languages. Tsou *tʔee* could reflect either form.

- (22) **PAn** *(t,C)ebuS ‘sugarcane’ > **Tso** tǎfsə **Kan** tevése **Saa** e-teve **RukMag** tbúsu **RukTon** tibóso (cf **RukTan** **RukBud** cobósə)

At first sight the change PAn *C > Tsou, Kanakanavu, Saaroa *t* in (20), (21) and (22) seems to be an innovation that defines the Tsouic subgroup. It is unlikely to have happened independently and thus satisfies the ‘no independent innovation’ condition. The question then is, Does it also satisfy the ‘no copying’ condition? The answer appears to be ‘no’. Indeed there are reasonable grounds to infer that the *C > *t* change was copied from Maga or Tona Rukai (or their shared ancestor) into the Tsouic languages.

The Maga Rukai and Tona Rukai cognates in (20a), (20b), (21a) and perhaps (22)—but not (21b) or (21c)—also display the *C > *t* change, yet Tanan, Budai and Mantaaran Rukai all reflect PAn *C as expected *c*. There is no reasonable doubt that the Rukai dialects are descended from a single source, Proto Rukaic, and that Proto Rukaic reflected PAn *C as *c. It follows that Maga and Tona Rukai underwent sibilant dissimilation after the break-up of Proto Rukaic, and that, if sibilant dissimilation is unlikely to have happened independently in each language, then either Maga and Tona copied it from a Tsouic language or the Tsouic languages acquired it from Maga or Tona.

Two facts suggest that sibilant dissimilation was a rule in Pre-Maga-Tona. The first fact is that the etyma in (23) provide additional evidence of the Maga-Tona rule.⁵⁶

- (23) a. **PAn** *CaCaS ‘dye yam, *Dioscorea cirrhosa*’ > **RukTon** tatásə (**RukTan** **RukBud** cocásə **Pai** tsatsas ‘*Tetrapanax papyrifus*, leaves used to colour fingernails’)
 b. **PRuk** *ciŋirisi ‘sprout from stump’ (v) > **RukMag** u-tŋĩrsi **RukTon** wa-itiŋĩisi (cf **RukMan** o-ciŋiri?i)
 c. **PRuk** *calase > ‘pus head’ > **RukMag** tálsə **RukTon** talásə (cf **RukBud** cálasə **RukMan** cala?ə)

The second fact is that the *C > *t* change does not apply to the Rukai reflexes in (21), where the appropriate conditioning is absent.

Copying could have taken one of two forms. Either the individual etyma were copied or the rule was copied (§2.2.2.2). The fact that the change also applies in Kanakanavu and Saaroa to the three items in (21) and to one Tsou item in (21c) implies that the rule was copied but overapplied in the Tsouic languages, as one might expect from copying. A further piece of evidence in support of this hypothesis is shown in (24): here the

⁵⁶ The Proto Rukai etyma in (23b-c) reflect putative PAn *CiŋiRi(sS) ‘sprout from stump’ (v) and *CaLa(sS) ‘pus head’, but no corroborating evidence for these reconstructions is known to me.

sibilant dissimilation rule fails to apply, apparently because *S had been lost in this item before the rule was applied in Kanakanavu.

- (24) **PAn** *CuSuR ‘thread (needle)’ > **Kan** c-um-a-cuúru

The evidence presented above is fragmentary and further lexical investigation of the languages concerned may show the details to be incorrect. But sibilant dissimilation has been copied in one direction or another, and if it has been copied between Tsouic and Rukai varieties, there is no reason to claim that it has not been copied from one Tsouic variety into another. What looked like a strong piece of subgrouping evidence is not what it seemed.

5.2.3 Paragogic vowels

A second innovation, which embraces all Tsouic and Rukai varieties, is the addition of a paragogic vowel after a PAn final consonant (Tsuchida 1976:32-33, 61-62, 88-89). This vowel is identical to the vowel of the root-final syllable, except that when the latter reflects PAn *a, the paragogic vowel is -ə. Again there are indications that this rule was copied rather than inherited, because its application in Tsou shows that it happened there after certain Tsou-specific changes.

Stress in pre-Tsou was unpredictable, at least in consonant-final roots.⁵⁷ The syllables on either side of the stressed syllable were unstressed, as were further alternate syllables in either direction. Pre-Tsou unstressed vowels other than /a/ were deleted sometime in the history of Tsou, unless this would result in an impermissible consonant cluster (Tsuchida 1976:88-89, 265-267). Paragogic vowels were added before the deletion of unstressed vowels from root-final syllables. We know this from examples like (25) where the paragogic vowel matches the lost vowel of the pre-Tsou final syllable:

- (25) a. **PAn** *C<um>aŋis ‘<AV>weep’ > pre-Tsou *t<um>aŋis > **Tso** m-oŋsi
 b. **PAn** *q<um>añud ‘<AV>drift, float away’ > pre-Tsou *ʔ<um>aLuc > **Tso** ŋ-ohcu
 c. **PAn** *C<um>aqiS ‘<AV>sew’ > pre-Tsou *t<um>aʔis > **Tso** t<m>eʔsi
 d. **PAn** *qubiS ‘pubic hair’ > pre-Tsou *(ʔ)úbis > **Tso** fsi-fsi

⁵⁷ ‘Pre-Tsou’ is simply an earlier stage of Tsou. It is not ‘Proto Tsouic’ as reconstructed by Tsuchida (1976) or by Pejros (1994). Wolff (1991) also used pre-Tsou stress as an element in his reconstruction of PAn stress. For a critique of the latter see Blust (2009:547-551, 553-556).

Examples (25a), (25b) and (25c) also illustrate the change whereby PAn *a becomes Tsou *o* or, before *-i-* in circumstances not fully understood, *e* (Tsuchida 1976:273). This change is involved in a curious set of exceptions to the paragogic vowel rule, noted by Tsuchida (1976:89). The items in (26) behave as the paragogic vowel rule predicts: root-final-syllable *-a- gives rise to paragogic *-ə*.

- (26) a. **PAn** *quzaL ‘rain’ > pre-Tsou *(?)úcaL > **Tso** m-əchə
 b. **PAn** *beRas ‘husked rice’ > pre-Tsou *fǿras > **Tso** fǿrsə
 c. **PAn** *Sepat ‘4’ > pre-Tsou *sǿpat > **Tso** səptə

The items in (27) misbehave, however. In (27a)-(27d), final *-o* occurs instead of expected final *-ə*. Tsuchida attributes this to conditioning: it occurs after pre-Tsou final *-or, *-oh and *-oʔ, but the examples are so few that we cannot be sure of this.⁵⁸ At any rate, paragogic *-o* appears to reflect pre-Tsou root-final-syllable *-o-, and in (27e) paragogic *-i* apparently reflects pre-Tsou *-e-.

- (27) a. **PAn** *S<um>apaR ‘<AV>lay a mat’ > pre-Tsou *s<um>ápar > **Tso** s<m>opro
 (**Kan** s<um>apárə **Saa** um-a-aparə)
 b. **PAn** *SimaR ‘fat, grease’ > pre-Tsou *simar > **Tso** simro (**Saa** ʔimarə)
 c. Pre-Tsou *t<um>aʔ-púrah ‘<AV>fall from cliff’ > **Tso** t<m>aʔ-purho (cf **Tso** s<m>o-prohə ‘<AV>jump down’)⁵⁹
 d. Pre-Tsou *ŋávaʔ > **Tso** ŋovʔo ‘<AV>hook’ (cf **Tso** m-ro-ŋvoʔə ‘slip off hook’)⁵⁹
 e. **PAn** *paRajaS ‘nettle tree, *Laportea* sp.’⁶⁰ > pre-Tsou *paráyas > **Tso** fresi (**Kan** parárasə **Saa** pararasə)

The critical point here is that the paragogic vowels in (27) reflect vowels that occurred in pre-Tsou or later, and their addition cannot have occurred in an ancestor common to Tsou, Kanakanavu and Saaroa. Instead they point to pre-Tsou rule copying of the kind discussed in §2.2.2.2.

⁵⁸ Tsuchida couches this in terms of synchronic morphophonemics whereas I have interpreted his statement in diachronic terms.

⁵⁹ Tsuchida (1976:89) cites no cognates for (27c) or (27d) but provides the forms in parentheses as evidence of deleted *-o-.

⁶⁰ This is my reconstruction, also taking account of Proto Rukaic *paʔagase, Pazih *baxasa*.

5.2.4 Reflexes of PAn *-an

Mentioned by Tsuchida (1976:33, 216-218) in conjunction with paragogic vowels is an innovation whereby Proto Tsouic unstressed final *-an (< PAn *-an) becomes Tsou, Kanakanavu and Saaroa *-a* rather than expected *-anə*. He infers the change sequence *-an > *-ã > *-a*. He notes (1976:11) that the same change occurs in Maga and Mantaaran Rukai. An examination of the relevant data suggests, however, that this is not a shared innovation but another instance of rule copying.

My analysis of the Tsou, Kanakanavu and Saaroa reflexes of PAn *-an differs somewhat from Tsuchida's. The Tsou change he posits presupposes a reconstruction of Proto Tsouic stress. But even if one accepted the existence of Proto Tsouic, it simply wouldn't be possible to reconstruct its stress. Tsuchida recognized that there were discrepancies between pre-Tsou and Kanakanavu stress⁶¹ and in such cases wrote an acute accent over the vowels of **both** possibly stressed Proto Tsouic syllables (Tsuchida 1976:321-329)! Pejros (1994) makes strong claims on the basis of reconstructed Proto Tsouic stress, but the number of instances where pre-Tsou and Kanakanavu disagree—over 25% (Pejros ignores them)—is enough to suggest that the relationship between pre-Tsou and Kanakanavu stress is either random or mediated by factors that are not understood. For pre-Tsou, on the other hand, stress is reconstructable as described in §5.2.3, and is shown in the pre-Tsou reconstructions in (28)-(32), provided to elucidate what has happened in Tsou.

In the items in (28) PAn *-an 'nominalizer' is reflected as *-a* in Tsou, Kanakanavu and Saaroa. Note, however, that not all three languages are represented in each item.⁶²

- (28) a. **PAn** *taLaS-an 'village' > **Tso** *hos-a* (< pre-Tsou *tahás-an) **Saa** *tala-a* **Tagalog** *tahán-an* 'home' (metathesis) [also **PAn** *taLaS > **Kan** *tanásə* 'village', **Tagalog** *tahan* 'reside']
- b. **PAn** *ka-pitu-an 'seventh month' > **Tso** *ʔ-pitv-a* (< pre-Tsou *ʔa-pítu-an) **Kan** *ka-pítu-a* **Saa** *ka-pitu-a* **Pai** *ka-pitju-an*
- c. **PAn** *paliSi-an 'taboo' > **Tso** *prisi-a* (< pre-Tsou *palísi-an/*palisi-an) **Saa** *palisi-a* (*-s-* for expected loss) **Tha** *parfi-an* **Sai** *pifi-an* [**Tamalakaw** **Puyuma** *palisi* 'pray']
- d. **PAn** *[ka-]La(?)uman-an 'when?' > **Tso** =*homn-a* (< pre-Tsou *háuman-an) **Kan** *-ka-naumán-a*, **PRuk** *ka-laʔum-ane (haplology: **RukMan** *ka-*

⁶¹ Saaroa is irrelevant here as it lacks phonemic stress.

⁶² Where there are items which reflect the root without *-an, these are placed in square brackets at the end of the set. In (28d) Saaroa *-taumanə* reflects PAn *La(?)uman, i.e. the root without the suffix.

laʔuma-a=ɔa ‘when (past)’, *ka-laʔuman-a=ɔa* ‘when (future)?’ [PAN *La(ʔ)uman ‘when?’ > **Saa** *-laumanʔ* **Bun** *lakua*]

In (29) Kanakanavu *caʔani-a* behaves like the items in (28), but Tsou *cohzon-a* is idiosyncratic. It appears that PAN *-an has been reanalyzed as part of the stem, and a second *-an has been added.

- (29) **PAN** *DaqaLi-an ‘daylight’ > **Tso** *cohzon-a* (< pre-Tsou *coohi-án-an) **Kan** *caʔani-a* **Pai** (western dialect) *rɔmʔaqati-ati-an* ‘mid-morning’ **Bab** *dalen-, lalian* (assimilation) **Paz** *dalian* ‘noon’ *ʔlan* ‘day’ [PAN *DaqaLi[an] ‘daylight’ > **Tha** *saqaði* (*s-* for expected *t-*) **Roviana** *rane*]

The items in (30) also reflect PAN *-an ‘nominalizer’. The reflexes in Tsou and Saaroa are *-a*, as expected, but the Kanakanavu reflex is *-anə*. Saaroa *ʔamis-an-a* (30a) and *taruan-a* (30b) are idiosyncratic in reflecting a double suffix, i.e. *qamiS-an-an and *taRuq-an-an (cf Tsou *cohzon-a* in (29)). Again this is apparently a reanalysis of the first *-an as part of the stem, perhaps under the influence of the Kanakanavu *-anə* forms.

- (30) a. **PAN** *qamiS-an ‘winter’ > **Tso** *ami-a* ‘north’ (< pre-Tsou *amí-an) **Kan** *ʔamis-anə* **Saa** *ʔamis-an-a* **Bun** *qamis-an* ‘year’ **Bab** *hamis-an* ‘north’ **Sai** *amif-an* **Paz** *amis-an* ‘north’ **PAta** *qamis-an **Tagalog** *amih-an* ‘northeast wind’ [**Puy** *ʔami* ‘year’ **Ami** *ka-qamis* ‘north’ **Bas** *amis* ‘north, west’ **Kav** *imis* ‘north’]
 b. **PAN** *taRuq-an ‘hunting hut’ > **Tso** *trova* (< pre-Tsou *tarú-an) **Kan** *tarúʔ-anə* **Saa** *taru-an-a* **PRuk** *taʔuanane (**RukMag** *tovnáa* **RukMan** *taʔun-aaə*) **Bun** *taluhan* **Tha** *taluyan* ‘field hut’ **Paz** *taxuan* **Sai** *ta[æʔæn* **Ami** *taluyan* **Aklanon** *taguyan* ‘hiding place’
 c. **PAN** *asik-an ‘broom’ > **Kan** *asik-anə* [PAN *asik ‘sweep’ > **Kan** *m-aru-asíki* ‘sweep’ **Bun** *ʔasik* ‘broom’ **Ami** *mi-ʔasik*]
 d. **PAN** *qajiS-an ‘boundary’ > **Tso** *es-a* (< pre-Tsou *(ʔ)áis-an) **Kan** *ʔáis-anə* **Saa** *ais-a* ‘between, middle’ **RukMag** *agís-nə* [PAN *qajiS ‘boundary’ > **PRuk** *agisi **Pai** *la-qedi-qedi* **Bun** *hais* **Sai** *æzis* **PAta** *qayis]

The items in (31) are roots that end in PAN *-an. Saaroa *ʔata* in (31a),⁶³ Tso *re-rpa* and Saaroa *ʔabalipa* in (31b) behave like the items in (28): *-an > *-a*. Kanakanavu *ʔabalipánə* in (31b) behaves like Saaroa *-laumanʔ* in (28d): *-an > *-anə*. Tsou *cronə*

⁶³ In (31a) Kanakanavu *ʔanái* is too irregular to analyze.

and Kanakanavu *caanə* in (31c) reflect *-an as *-onə/-anə*, i.e. like the Kanakanavu items in (30).

- (31) a. **PAn** *ŋajan ‘name’ > **Kan** *ŋandí* (for expected †*ŋala* or †*ŋalanə*) **Saa** *ŋala* **Puy** *ŋadan* **PRuk** *nagane (**RukBud** *nagánə*) **Pai** *ŋadan* **Bun** *ŋan* **Ami** *ŋaŋan* **PAta** *ŋahan (**Sed** *ŋaxan*) **Ami** *ŋaŋan* (assimilation) **Kav** *naŋan* (metathesis)
- b. **PAn** *[qa]lipan ‘centipede’ > **Tso** *re-rpa* (< pre-Tsou *lilipan) **Kan** *ʔalʔalipáŋə* **Saa** *ʔalʔalipa* **Paz** *haripan* **Mal** *halipan*
- c. **PAn** *zalan ‘road’ > **Tso** *cronə* (< pre-Tsou *colón) **Kan** *caánə* (for expected †*cala* or †*calanə*) **PRuk** *dalane **Puy** *daŋan* **Pai** *djalan* **Hoa** *dzalan* **Sir** *darəŋ* **Malay** *jalan*

Finally, the items in (32) are PAn roots ending in *-aŋ. The Tsou and Kanakanavu reflexes behave as expected: they add a paragogic vowel. The Saaroa reflexes, on the other hand, behave as if the roots ended in *-an and reflect *-aŋ as *-a*.

- (32) a. **PAn** *kurapaŋ ‘toad’ > **Saa** *ru-kurapa* **RukMan** *kurapaŋə* **Pai** *tja-kurapaŋ* ‘rain frog’
- b. **PAn** *qaLiwaDaŋ ‘shoulder’ > **Kan** *ʔaniʔacáŋə* **Saa** *aliasa* **Ilokano** *aliwadang* [**PAn** *qaLiwaDaŋ-an > **RukMan** *aliaðaŋ-aaə* **Puy** *qaliwaqang-an* ‘collar bone’]
- c. **PAn** (?)⁶⁴ *LuCaŋ ‘man (male)’ > **Tso** *ha-hocŋə* (< pre-Tsou *húcoŋ) **Saa** *ta-husa*

The patterns exemplified in (28)-(32) are summarized in (33). There is no inherited Proto Tsouic pattern here. Instead, each language displays a pattern of its own.

(33) PAn	Tsou	Kanakanavu	Saaroa	examples
*-an ‘nominalizer’	<i>-a</i>	<i>-a</i>	<i>-a</i>	(28)
*-an ‘nominalizer’	<i>-on-a</i>	<i>-a</i>	<i>-a</i>	(29)
*-an ‘nominalizer’	<i>-a</i>	<i>-anə</i>	<i>-a, -an-a</i>	(30)
root in *-an	<i>-aŋə</i>	(28d)
root in *-an	<i>-a</i>	<i>-aŋə</i>	<i>-a</i>	(31a-b)
root in *-an	<i>-onə</i>	<i>-anə</i>	...	(31c)
root in *-aŋ	<i>-(a)ŋə</i>	<i>-aŋə</i>	<i>-a</i>	(32)

⁶⁴ ‘PAn (?)’ indicates a possible PAn form that is not widely enough reflected to be firmly reconstructed.

Tsuchida's generalization that unstressed *-an is reduced to -a does apply to Tsou. If pre-Tsou stress is reconstructed on the basis of the retention or deletion of pre-Tsou vowels in Tsou, then it emerges that in just two cases, *cohzon-a* in (29) and *cronə* (31c), pre-Tsou stress fell on the syllable reflecting *-an, and *-n is retained in Tsou. In all other cases, including the second *-an of (29), pre-Tsou *-an was unstressed and becomes Tsou -a.

The only conditioning displayed by Kanakanavu is that in all roots ending in *-an or *-an̩, the final syllable is usually reflected as -an̩, i.e. *-n becomes -n̩, and then (as predicted) paragogic -ə is added. The one exception is *cadánə* 'path' in (31c), which is also irregular in other respects. A possible conditioning of the reflex of *-an 'nominalizer' is the number of syllables in the root: the suffix is -a with a trisyllabic root, -an̩ with a disyllabic root. On Pejros' (1994) analysis of Tsuchida's data, the stress in trisyllables falls on the penult of the stem (CVCVCV(C-v)), of disyllables on the stem-final syllable (CVCVC-v). This means that *-an is a syllable further away from the main stress in a trisyllable, and may account for its reduction to -a.

In Saaroa, final *-an and *-an̩ both become -a, regardless of whether *-an is a suffix or ends a stem. The only exception is *-laumanə* 'when?' in (28d), apparently reflecting PAn *La(?)uman, and I have no explanation for this.

The situation in the Rukai dialects is different again. In Li's (1977) Rukaic data the Proto Rukaic nominalizer *-ane is reflected as -ane in the Budai and Tona dialects, but as -aa in Tanan, -ae in Mantauran and either -i (unstressed) or -áa in Maga. The Tanan, Maga and Mantauran changes do not affect roots in *-ane, e.g. Proto Rukaic *-kane 'eat', *agane 'grandchild', *daʔane 'house'.

The conclusion to be drawn from (33) and Rukaic is that superficially similar but different processes have applied in the three languages and in the Tanan, Maga and Mantauran dialects of Rukai, where only suffix *-an is affected. This again bears the hallmark of contact, not of shared inheritance.

5.2.5 Tsuchida's mergers of coronal obstruents

Tsuchida (1976:253-254) lists a number of PAn mergers that he claims are reflected in all three Tsouic languages. Some are due to PAn distinctions (e.g. *L vs *N) which more recent work does not recognize. The others need to be considered in the same way as Blust's mergers in Table 5.

Two of these mergers concern coronal obstruents. To explicate them we need to look at more recent literature. Ross (1992) took it that PAn had the three phonemes *d₁, *d₂ and *d₃, reconstructed by Dahl (1973, 1976, 1981). Blust (1999) argued that their putative reflexes were the result of copying between Puyuma and Paiwan of etyma

reflecting PAn *d, but Ross (2011) shows that the patterning of reflexes does support the reconstruction of *d₁ and *d₂ and, less strongly, of *d₃.

PAn *d₁, *d₂ and *d₃ are relabelled *d, *Z and *D respectively by Ross (2011) (Blust had already relabelled Dyen's and Tsuchida's *Z as *z). The criterial reflexes of PAn *z, *d, *Z, *D and the corresponding reflexes of *C are shown in Table 9,⁶⁵ along with Dahl's and Tsuchida's notations.

Table 9: Formosan reflexes differentiating two or more of PAn *z, *d, *Z and *D, together with reflexes of PAn *C

PAn (Dahl, Ross 1992)	*Z	*d ₁	*d ₂	*d ₃	*C
Tsuchida (1976)	*Z	*D ₃	*D ₂	*D ₄	*C
PAn (Ross 2011)	*z	*d	*Z	*D	*C
Puyuma	d	d	d/z	d/z	t
Proto Rukaic	*d	*D	*D	*D	*c
Tsou	c	c	c	c	c
Kanakanavu	c	c	c	c (?)	c
Saaroa	s	s	s	s	c
Paiwan	dj	dj	z	d	ts
Thao	s	s	s	t	θ
Favorlang-Babuza	t	t	d, r	r	tʃ
Hoanya	dz	d	d	...	s
Siraya	d, l	s	s	s	t
Amis	l	r, l	r, l	r	t
PMP	*Z	*d	*d	*d	*t

5.2.5.1 Merger of PAn *C and Tsuchida's PAn *d

Tsuchida's *d does not appear in the second row of Table 9 as it appears not to be reconstructable. If this is true, then there was no *C/*d merger.

In Tsuchida's data the alleged merger occurs in just the cognate sets in (34) (Tsuchida 1976:181-182, 222; the reconstructions are mine), and is distinguished from his PAn *D₃ (my *d) by its Saaroa reflex, *c* rather than *s*. Saaroa *c* otherwise reflects PAn *C. In Ross (2011) the initial consonant of (34a), PAn *daRaq, is reconstructed on the basis of Puyuma *d*- and Paiwan *dj*-, and the final consonant of (34b), PAn *lahuZ, on the basis of Puyuma *-d* and Paiwan *-z*. In (34c) *zeŋeR is tentatively reconstructed with *z- on the evidence of Proto Rukaic *d- (there are no non-Formosan reflexes). The reflexes of the coronals in (34d) and (34e) are ambiguous. In (34d) both the initial and

⁶⁵ PAn *z, *d, *Z and *D merge as Bunun *d*, Pazih *d*, Saisiyat *r*, Proto Atayalic *d*, Kavalan *z*, and are accordingly not shown in Table 6.

final correspondences are irregular, and it is possible that Saaroa *-cakici* (for expected \dagger -*takisi*) reflects a conflation with the expected Saaroa reflex of PAN **dekeC* ‘adhere’ (Puyuma *deket* Paiwan *djekets* PMP **deket*), namely \dagger *sakici*.

- (34) a. **PAn** **daRaq* ‘blood’ > **Kan** *caráʔə* **Saa** *caraʔə* **Puy** *daraʔ* **Pai** *djaq* **Bab** *tagga* **Hoa** *ddoi* **Malay** *darah*
 b. **PAn** **lahuZ* ‘seaward’ > **Tso** *moh-rovcu* ‘flow downhill’ **Kan** *ʔama-laúcu* ‘downhill’ **Saa** *tala-la-laucu* ‘look down’ **Puy** *lauq* ‘east’ **Pai** *lauz* ‘seaward’ **PRuk** **LauDu* ‘downhill’ **Sir** *r<m>aus* **Malay** *laut* ‘sea’
 c. **PAn** (?) **zeŋeR* ‘discharge from ear’ > **Kan** *pi-ceŋe-céŋer-a* **Saa** *pi-ceŋe-ceŋer-a* **PRuk** **deŋe-deŋe-ane*
 d. **PAn** **taki(d,Z,D)* ‘adhere’ > **Kan** *m-aa-takici* **Saa** *mai-ca-cakici* **Kapampangan** *takid*
 e. **PAn** **(z,d)amay* ‘side dish’ > **Kan** *camái* **Saa** *camai* **Pai** *djamay* **Tha** *samað* (for *samay*)

The sets in (35) also have Saaroa *c* where *s* is expected, but these sets are also irregular in other respects.

- (35) a. **PAn** **(z,d,Z)aqu* ‘the soapberry, *Dracontomelon edule*’ > **Kan** *Saa* *caaʔu*, **Puy** *daʔu* **Rukai** *daw* (Wolff 1997, dialect not identified) **Pai** *zaqu* **Ami** *rauʔ* **Malay** *rau*
 b. **PAn** (?) **(z,D,Z)aLu(C,q)* ‘transplant rice seedlings’ > **Tso** *t<m>uhcu* **Kan** *c<um>a-canúcu* **Saa** *c<um>a-calucu* **Bun** *ma-danuq* (-*q* for expected -*t*)

The tidiest explanation of the apparently random distribution of Saaroa *c* across sets reflecting PAN **z*, **d* and **Z*, and perhaps **D*, is that Saaroa *s* regularly reflected all four PAN voiced coronals and Saaroa *c* regularly reflected PAN **C*. The reconstruction of Tsuchida’s **d* would require that the etyma in (34) display a regular sound correspondence which includes Saaroa *c*, but this requirement is not satisfied, indicating that Tsuchida’s **d* did not exist and that its alleged reflexes reflect other coronals. His **C*/**d* merger thus did not occur, and instead the occurrences of Saaroa *c* where *s* is expected imply lexical copying, an inference supported by the further irregularities in (35).

5.2.5.2 Merger of PAN **z*, **d*, **Z* and **D*

Tsuchida’s second coronal merger is of his **Z*/**D*₃/**D*₂/**D*₄, or **z*/**d*/**Z*/**D* in my 2011 notation (Table 9). The examples in (36), (37) and (38) confirm the reflexes of PAN **z*, **d* and **Z* shown in Table 9, namely Tsou *c*, Kananavu *c* and Saaroa *s*, with just two

exceptions, in (37b) and (38e), where Saaroa has *c*, discussed as (34a) and (34b) above.

- (36) a. **PAn** *zalan ‘road’ > **Tso** *cronə* **Kan** *caanə* **Saa** *m-u-salan-a* ‘walk (AV IMP)’ **PRuk** *dalane **Puy** *daʃan* **Pai** *djalan* **Hoa** *dzalan* **Sir** *daraŋ* **Malay** *jalan*
- b. **PAn** *zaLiH ‘near’ > **Kan** *ará-cani* **Saa** *ma-sati* **RukMag** *me-d-dali* **Pai** *djati* ‘soon’, **Sed** *dalix* **Ilokano** *dan-dani*
- c. **PAn** *zawiL ‘far’ > **Tso** *covhi* **Kan** *ara-caini* **Saa** *ma-saila* **Puy** *dawil* **PRuk** *daili **Tao** *tavit*
- d. **PAn** *quzaL ‘rain’ > **Tso** *m-əchə* **Kan** *ʔucánə* **Saa** *usatə* **Puy** *ʔudal* **PRuk** *udale **Pai** *qudjaʔ* **Bun** *qudan* **Tha** *qusað* **Hoa** *m-udzəs* **Sir** *udal* **Paz** *ʔudal* **Sai** *ʔæ-ʔæral* **Ami** *ʔurad* **Kav** *uzan* **Malay** *hujan*
- (37) a. **PAn** *dapaL ‘sole’ > **Tso** *caphə* ‘foot’ **Saa** *sapaʔə* **Puy** *dapal* **PRuk** *Dapale **Pai** *djapat* ‘hind legs’ [**Sir** *sapal*]
- b. **PAn** *daRaQ ‘blood’ > **Kan** *caráʔə* **Saa** *caraʔə* **Puy** *daraʔ* **Pai** *djaq* **Bab** *tagga* **Hoa** *ddoi* **Malay** *darah*
- c. **PAn** *qañud ‘drift’ > **Tso** *ŋ-óhcu* **Kan** *m-a-ʔacínu* (metathesis < pre-Kan *ʔalud) **Saa** *m-u-alusu* **Puy** *laʔud* (metathesis < pre-Puy *ʔalud) **PRuk** *aluDu **Pai** *qatudj*
- d. **PAn** *dilaq ‘lick, kiss’ > **Kan** *c(um)á-cəʔə* **Saa** *s(um)a-silaə* **PRuk** *Dila **Pai** *dj(um)ilaq* **Bab** *ta-tsira* (*ts* = affrication before /i/)
- e. **PAn** *daReq ‘soil, clay’ > **Tso** *crówa* **Saa** *sarəə* **Puy** *dareʔ* **PRuk** *daʔe (*d- for †*D-) **Fav** *ta* **Hoa** *rubrul-daxu* **Roviana** *raro* ‘clay pot’
- (38) a. **PAn** *ZaLum ‘water’ > **Tso** *chumu* **Kan** *canumu* **Saa** *salumu* **PPuy** *qanum **Pai** *zahum* **Tha** *saðum* **Bab** *dalom* **Sir** *salum* **Kav** *zanum* **Sir** *salum* **Ngaju** **Dayak** *danum*
- b. **PAn** *Zaya ‘upriver, inland’ > **Kan** *m-a-a-cála* ‘blow towards mountainside’ **Puy** *qaya* ‘west’ **PRuk** *Daða **Pai** *i-zaya* **Sir** *taxa-seya* **Kav** *zaya* ‘west’ **Malay** *barat-daya* ‘southeast’
- c. **PAn** *ZemZem ‘dark’ > **Tso** *cməcmə* ‘rainclouds’ **Kan** *m-u-a-cəməcəmə* ‘leave early in the morning’ **Saa** *səəsəm-a* **Puy** *qeme-kerem* ‘dawn twilight’ **Pai** *zemzem* **Bun** *dumdum* **Hoa** *ma-dum* **Sir** *ma-simdim* **Kav** *mRi-zemzem* ‘be dim’ **Itbayaten** *mir-remden* ‘become cloudy’
- d. **PAn** *Zuma ‘other’ > **Tso** *mo-cmo* **Kan** *ca-cumá-ini* **Saa** *u-suma-anə* **Puy** *quma* **PRuk** *Duma **Pai** *zuma* **Fav** *roman* **Kav** *zuma* **Ilokano** *ag-duma* ‘differ’

- e. **PAn** *lahuZ ‘seaward’ > **Tso** *moh-rovcu* ‘flow downhill’ **Kan** *ʔama-laúcu* ‘downhill’ **Saa** *tala-la-laucu* ‘look down’ **Puy** *[auq]* ‘east’ **PRuk** *LauDu ‘downhill’ **Pai** *lauz* ‘seaward’ **Sir** *r(m)aus* **Mal** *laut* ‘sea’
- f. **PAn** *layaZ ‘*Sambucus formosana*’ > **Tso** *hzocə* **Kan** *nalacə* **Puy** *layaq* **PRuk** *laðaD **Pai** *layaz* **Kav** *layas*
- g. **PAn** *likuZ ‘back’ > **Tso** *mi-riʔcu* ‘look back’ **Kan** *tara-ʔiku-ʔikúcu* ‘look back’ **Saa** *ʔilikusu* ‘back (body part)’ **Puy** *[ikuq-an]* ‘back (direction)’ **PRuk** *rikuDu ‘back (direction)’ **Pai** *likuz* **Sir** *rikus* **Kav** *ku-rikuz* ‘follow’ **Ngaju Dayak** *ba-rikor* ‘on one’s back’
- h. **PAn** *quZip ‘alive’ > **Kan** *ʔum)a-ʔuci-ʔucipi* **Saa** *maŋ-usipi* **PRuk** *m-uDipi **Pai** *q(m)uzi-quzip* **Fav** *orich*

Did PAn *D also merge with *z/*d/*Z as Tsou *c*, Kanakanavu *c* and Saaroa *s*? Quite possibly, as (39a) and (39b) indicate. However, reflexes are few and, as the expected forms shown in parentheses indicate, both (39d) and (39e) have irregular Tsou reflexes.⁶⁶ In (39e) the Kanakanavu reflex is also irregular. The expected reflex is †ʔucásə, and it is possible that Kanakanavu *ʔusásə* is the result of assimilation of *c* to *s*.

- (39) a. **PAn** *Deme[IR] ‘thick (as of a board)’ > **Tso** *o-cmərə* **Kan** *maki-cemere* **Saa** *maki-semere* **Puy** *ke-zemer* **PRuk** *Demele **Pai** *ke-qemel*
- b. **PAn** *DaqaLi[an] ‘daylight’ > **Tso** *cohzona* **Tha** *saqəði* (*s-* for expected *t-*) **Bab** *dalen-*, *lalian* (assimilation) **Pai** (western dialect) *r(m)aqəti-ati-an* ‘mid-morning’ **Paz** *dalian* **Chamorro** *haʔani* **Roviana** *rane*
- c. **PAn** *Dakes ‘camphor laurel, *Cinnamomum camphora*’ > **Tso** *cʔosə* **Kan** *cakése* **Puy** *qakes* **PRuk** *Dakese **Pai** *Dakes* **Tha** *fakif* (*f-* for expected *t-*) **Bun** *dakus* **Paz** *dakes* **Sai** *rakef* **PAta** *dakus **Ami** *rakes* **Kav** *zaqes*
- d. **PAn** *DuSa ‘two’ > **Tso** *ruso* (*r-* for expected *c-*) **Kan** *cuiśa* **Saa** *u-sua* **Puy** *q̄ua* **Pai** *q̄usa* **PRuk** *Dusa **Bun** *dusa* **Tha** *tufa* **Tao** *gua* **Bab** *na-doa*, *na-roa*⁶⁷ **Hoa** *rusa*⁶⁸ **Sai** *rufa* **Paz** *dusa* **PAta** *dusa **Sir** *duha* **Bas** *lusa* **Kav** *zusa* **Malay** *dua*
- e. **PAn** *quDaS ‘grey hair’ > **Tso** *ɲtosə* (for expected †(ʔ)cusə) **Kan** *ʔusásə* (for expected †ʔucásə) **Saa** *ʔusaə* **PRuk** *uDasə **Pai** *quqas* **Bun** *hudas* **Tha** *qutaf* **Fav** *uras* **PAta** *qudas

⁶⁶ Tsuchida accounts for the irregularity of Tsou *rusa* ‘two’ by reconstructing a separate protophoneme, reflected only in this root.

⁶⁷ PAn *S is irregularly lost in Favorlang-Babuza, as it is also in closely related Taokas *gua*.

⁶⁸ This is the only Hoanya reflex of PAn *D. Initial *r-* here may represent the same phoneme as *d* reflecting *d and *Z.

These cognate sets indicate that a merger of **z*, **d*, **Z* and **D* is indeed reflected in Tsou, Kanakanavu and Saaroa. However, parallel mergers are reflected as Bunun *d*, Thao *s*, Pazih *d*, Saisiyat *r*, Proto Atayalic **d* and Kavalan *z*, so the merger does not in itself satisfy the ‘no independent innovation’ condition.

One might argue, however, that the outcomes of the merger of **z*, **d*, **Z* and **D*, namely Tsou *c* ([tʃ]/_i; [ts] elsewhere), Kanakanavu *c* ([tʃ]/_i; [cç] elsewhere) and Saaroa *s*, are sufficiently unusual to suggest that the merger is indeed a shared Proto Tsouic innovation. Saaroa *s* would be accounted for by a change **c* ([ts] > *s*). But the process was not as straightforward as this. Tsuchida reconstructs the relevant developments as in (40):

(40)	PAn		‘Proto Tsouic’		Tsou	Kanakanavu	Saaroa
	<i>*C</i>	>	Tsuchida’s <i>*c</i>	>	<i>c</i>	<i>c</i>	<i>c</i>
	<i>*z/*d/*Z/*D</i>	>	Tsuchida’s <i>*č</i>	>	<i>c</i>	<i>c</i>	<i>s</i>

Tsuchida (1976:253) recognizes that, as Saaroa reflects **C* and **z/*d/*Z/*D* differently, his Proto Tsouic must also have kept them apart. They have merged in Tsou and Kanakanavu. By any criterion Kanakanavu is more closely related to Saaroa than it is to Tsou, and so the Tsou and Kanakanavu mergers must have been separate. It is probable that PAn **C* was phonetically [ts] and has more or less maintained this articulation in Tsou, Kanakanavu and Saaroa, as well as in Rukai and Paiwan. Whatever the pre-Saaroa outcome of the **z/*d/*Z/*D* merger, it wasn’t [ts], and a [ts] > [s] origin for Saaroa *s* is unsustainable.

5.2.6 Other mergers

The second phoneme in each of the remaining mergers posited by Tsuchida—**k/*g*, **R/*r*, **S/*θ*—has such a low functional load that we do not have enough relevant data to evaluate the mergers thoroughly.

5.2.6.1 Merger of PAn **k* and **g*

The **k/*g* merger is difficult because **g* is reflected convincingly in only the few items in (42) and because firm reflexes of **g* in a number of Formosan languages are unavailable. However, it appears that PAn **k* and **g* had indeed merged in Tsou, Kanakanavu and Saaroa—and also in Proto Rukaic, Bunun, Pazih and Proto Atayal, thereby implying the possibility of independent parallel innovation. As (41) and (42) show, Tsou has two reflexes of each, *ʔ* and *k*, of which *ʔ* is the more frequent, and presumably the inherited reflex.

- (41) a. **PAn** *kuCu ‘head louse’ > **Tso** *ʔcuu* **Kan** *kúucu* **Saa** *kucuʔu* **Puy** *kufu*
PRuk *kucu **Pai** *kutsu* **Bun** *kutu* **Tha** *kuθu* **Paz** *kusu* **Sai** *kusu* **PAta**
kucu* **Kav *qutu*
- b. **PAn** *kaRaC ‘bite’ > **Tso** *b-orcə*,⁶⁹ *ro-ʔroc-a* **Saa** *um-a-aracə* **Puy**
k(um)arəʔ **PRuk** *ua-kaʔace **Pai** *k(em)ats* **Bun** *kalat* **Sai** *k(um)aləs* **PAta**
k(um)agac* **Kav *q(m)aRat*
- c. **PAn** *kaen ‘eat’ > **Tso** *b-onə*,⁶⁹ *an-a* **Saa** *um-an-a*, *a-anə* **Puy** *me* **Kan**
PRuk *ua-Kane **Pai** *k(em)an* **Bun** *kaun* **PAta** *k(um)an **Kav** *q(m)an*
- d. **PAn** *Dakes ‘Camphor laurel’ > **Tso** *cʔosə* **Kan** *cakəsə* **Puy** *dakes*
PRuk *Dakese **Pai** *dakes* **Bun** *dakus* **Tha** *fakif* (ʃ- for expected s- or t-)
Paz *dakes* **Sai** *rakeʃ* **PAta** *dakus **Kav** *zaqes*
- e. **PAn** *bukeS ‘head hair’ > **Tso** *fʔəsə* **Kan** *vəkəsə* **Saa** *vəkəə* **Tha** *fukif*
Paz *bekes* **Sai** *bukeʃ* **Kav** *vukes* **Tagalog** *buhok* (*S metathesis)
- f. **PAn** *kaRaw ‘scratch’ > **Tso** *tuo-krov-a* **Kan** *k(um)a-karə-kara* **PRuk**
ka-kaʔaw* **Bun *ma-kalav* **Paz** *kaxaw* **Sai** *k(um)a-kaʔaw* **Kav** *qaRaw*
- g. **PAn** *(r,l,L)utuk ‘rabbit’ > **Tso** *rotúka* **Kan** *Saa* *lituka* **PRuk** *Lutuku
Pai *rucuk* **Sai** *ʔutuk*
- (42) a. **PAn** *gerger ‘shake, shiver, tremble’ > **Tso** *s(m)o-ʔerʔere* **Kan** *maa-*
kerekere **Saa** *k(um)a-kekere* **Pai** *mi-gerger* **Bun** *ku-kulkul* **Ilokano** *pi-*
gergér
- b. **PAn** *tageRaŋ ‘ribs’ > **Tso** *tʔorŋa* ‘chest’ **Kan** *takérəŋa* ‘sternum’ **Puy**
tageraŋ ‘chest’ **PRuk** *tokoʔoso **Paz** *takaxaŋ* **Sed** *tekelaŋ* **Pangasinan**
taglaŋ
- c. **PAn** *guluŋ ‘roll up’ > **Tso** *t(m)ai-ku-kruŋu* ‘(wheel) roll’ **Tagalog**
g(um)úləŋ ‘roll’

5.2.6.2 Merger of PAn *R and *r

The *R/*r merger is problematic because few PAn reconstructions contain *r. The few of these that are reflected in Tsou, Kanakanavu and Saaroa have merged *R and *r as Tsou *∅* before a consonant and *r* elsewhere and as Kanakanavu and Saaroa *r*, but only one Formosan language, Paiwan reflects them separately (*r as *r*, but *R is lost). Thus the merger does not satisfy the ‘no independent innovation’ condition.

⁶⁹ Tso *b-orcə* and *b-onə* reflect pre-Tsou *ʔ(m)orcə and *ʔ(m)onə respectively.

5.2.7 Changes in sibilants

Neither *θ nor *x (Tsuchida's *S₂) is reconstructed by Blust, but Ross (2011) argues that they should be reconstructed,⁷⁰ even though neither occurs in many PAn etyma. Their reflexes in Rukai (the only language to retain a separate reflex of *θ), Tsou, Kanakanavu and Saaroa are shown in (43). Tsuchida posits two Tsouic sibilant innovations, the loss of *x and a merger of his *S₆ and *θ.

(43)	PAn	*s	*θ	*x	*S
	Proto Rukaic	*s	*θ-θ-s	*s	*S
	Tsou	s	s	∅	s
	Kanakanavu	θ-θ-s/θ	s	∅	s-s/θ-s
	Saaroa	∅	s	∅	s/θ-θ-s/θ ⁷¹

The loss of *x does not meet the 'no independent innovation' condition, as it is also lost in Puyuma, Thao, Taokas, Favorlang-Babuza, Papora and Hoanya.

The *S₆/*θ merger (Tsuchida 1976:253)⁷² turns out to be like the merger of *C and Tsuchida's *d (§5.2.5.1). It involves a protophoneme that is not reconstructable. His *S₆ is a variant of *S. *S is, on Tsuchida's analysis (1976:159), normally lost in Saaroa, but *S₆ represents its retention as *s* in certain items. Thus for Tsuchida *S₆ and *θ merge as Tsou, Kanakanavu and Saaroa *s*. I infer instead that *S has been lost sporadically in Saaroa (or was lost categorically and reintroduced through contact with closely related Kanakanavu), and that there is no reason to reconstruct *S₆. If *S₆ is not reconstructed, then there was no *S/*θ merger.

A glance at (43) might also suggest a shared innovation whereby Tsou, Kanakanavu and Saaroa all reflect *θ as *s*, but the behaviour of Tsou in relation to the sibilants is distinct from that of Kanakanavu and Saaroa. Tsou merges PAn *s, *θ and *S as *s*, whereas Kanakanavu and Saaroa reflect the three phonemes differently. If we assume the worldwide diachronic changes [s] > [h] > ∅, then Kanakanavu and Saaroa share the fact of having taken this process furthest with *x and *s (with Kanakanavu lagging a little behind in its lenition of *s), less far with *S, and nowhere at all with *θ. There is

⁷⁰ The argument is that, unlike a number of Tsuchida's subscripted protophonemes, *θ and *x are each supported by a **set** of distinct reflexes: *θ as shown in (43) and *x by Proto Rukaic *s, Paiwan, Bunun, Amis and Kavalan *s*, Pazih, Saisiyat *h*, Proto Atayalic *h, and loss in Puyuma, Tsou, Kanakanavu, Saaroa, Thao, Taokas, Favorlang-Babuza, Papora and Hoanya.

⁷¹ *S is retained root-medially in just one etymon, PAn *paLiSi 'taboo', Saaroa *palisi-a*.

⁷² Tsuchida writes of a merger between *S and *θ, but there is a misprint here: the merger is between his *S₆ and *θ.

thus evidence here that Kanakanavu and Saaroa have shared a period of common development.

5.2.8 Interim conclusions

In (44) the innovations which allegedly define the Tsouic subgroup are listed together with reasons for not according them subgroup-defining status and a reference to the subsection in which each is discussed.

(44)	Innovation	evaluation	crossreference
	sibilant dissimilation	fails ‘no copying’ condition	§5.2.2
	paragogic vowels	fails ‘no copying’ condition	§5.2.3
	*-an > -a	fails ‘no copying’ condition	§5.2.4
	*C/*d merger	Tsuchida’s *d did not occur	§5.2.5.1
	*z/*d/*Z/*D merger	fails ‘no independent innovation’ condition	§5.2.5.2
	*z/*d/*Z/*D > c	no Saaroa support	§5.2.5.2
	*k/*g merger	fails ‘no independent innovation’ condition	§5.2.6.1
	*R/*r merger	fails ‘no independent innovation’ condition	§5.2.6.2
	*S ₆ /*θ merger	Tsuchida’s *S ₆ did not occur	§5.2.7
	loss of *x	fails ‘no independent innovation’ condition	§5.2.7

The shared phonological innovations in (44) which fail the ‘no independent innovation’ condition do so quite spectacularly, as each is reflected in several other Formosan languages.

None of the innovations in (44) stands up to inspection as subgroup-defining. Instead the first three innovations are evidence of contact, not only among Tsou, Kanakanavu and Saaroa but also with the ‘Lower Three Villages’ dialects of Rukai (Maga, Tona, Mantauran). If we believe the evidence of the single example in (24), then sibilant dissimilation did not occur in Kanakanavu until loss of medial *-S- in Kanakanavu had begun. Paragogic vowels were evidently added in Tsou before unstressed syllables were lost. Further evidence of contact is clearly provided by the close lexical similarity among Tsou, Kanakanavu and Saaroa and the slightly less close lexical similarity between these three languages and Rukai.

The picture that emerges here is one of languages that are related, but no longer closely related, coming back into closer contact with each other after a period of separation so that innovations again pass between them. Within Austronesian, this is similar to the processes noted by Geraghty (1983:379-386) for the Western Fijian and Tokalau Fijian linkages and by Ross (1997:230) for New Ireland, where separate

linkages were reintegrated through contact between geographically proximate languages of the two linkages.

5.3 A Rukai-Tsouic subgroup?

These interim conclusions lead us back to the question raised by Tsuchida (1976: 10-11): is there a Rukai-Tsouic subgroup? The answer is ‘no’. If there is no Tsouic subgroup, as I have claimed in §5.2.8, there can be no Rukai-Tsouic group. There is also no evidence to suggest that Rukai subgroups more closely with one or more of Tsou, Kanakanavu and Saaroa than with any other Formosan language. Instead, it follows from the first three innovations in (44) that the shared features of the four languages are the outcome of contact, not of shared inheritance. Blust comments,

The ‘echo vowels’ of Tsouic and Rukai, as well as the preglottalized consonants of Thao, Bunun, and Tsouic appear to be fairly clear-cut cases of the diffusion of phonetic features within restricted portions of Taiwan. (Blust 1999:33)

and further:

As noted by Tsuchida (1976), they [the Rukai dialects] share some features with the Tsouic languages, as the addition of supporting vowels, and the loss of final *-n after *a. Since the latter change is confined to the Maga and Mantauran dialects I regard it, like the common development of supporting vowels in all Rukai dialects, as a product of contact. (Blust 1999:51-52)

Tsuchida (1976:11-12) lists a sample of words that are uniquely reflected in Rukai and one or more of Tsou, Kanakanavu and Saaroa. If the NAn hypothesis is correct, then there are two ways in which this shared lexicon might have come into being. One is, of course, copying. The second is that, since Rukai, Tsou and the pair Kanakanavu and Saaroa belong to three different subgroups, some such items are retentions from PAN that have been lost elsewhere in Austronesian but have persisted because of ongoing contact.

6. Conclusions

The goal of this paper has been to defend the Nuclear Austronesian hypothesis, whereby all Austronesian languages other than Puyuma, Tsou and Rukai belong to a single subgroup, Nuclear Austronesian (§3). This subgroup is defined by the nominalization-into-verb innovation. In the early part of the paper the nature of

innovations that establish a subgroup was examined, along with the likelihood that in their early stages the ancestors of the Formosan languages formed a dialect network (§2.1). Discussion of current hypotheses in §3 and §4 led to the conclusion that during the earliest reconstructable period of Austronesian, PAn split into four languages, the ancestors of Puyuma, Tsou, Rukai and NAn, and that the last diversified into a linkage which maintained its immediate interconnectedness longer in the north and west than in the east and south (§4.4).

The defense of NAn itself falls into two parts. First, other recent subgroupings of Austronesian languages were examined to see whether they stood up to close examination (§4). My conclusion is that in a number of respects they don't, but that Ho's and Blust's hypotheses include the adequately supported low-order subgroups Western Plains and Atayalic, and Ho's Atypical Formosan (Siraya, Amis, Kavalan, Bunun and PMP) has respectable support, in the sense that it is worthy of further investigation. I have also sought to explain the data on which Ho's, Blust's and Sagart's subgroupings are based, providing a reanalysis of Sagart's data (§4.3.2), as his hypothesis is the least compatible with Nuclear Austronesian.

Second, because the widely accepted Tsouic subgroup is incompatible with Nuclear Austronesian (since Tsou lies outside Nuclear Austronesian, Kanakanavu and Saaroa within it), I have tried to show how radically different Tsou is from its neighbors (§5.1) and made a detailed examination of the arguments that have sometimes been proposed as supporting the Tsouic subgroup, and have found that they are strongly suggestive of longterm contact but are not the kind of evidence required to establish a subgroup (§5.2). The inference is that Kanakanavu and Saaroa came into contact with Tsou after a period during which the histories of Tsou and Kanakanavu-Saaroa were quite different. The rejection of Tsouic leads automatically to a rejection of Rukai-Tsouic (§5.3).

It is a widely held principle of historical linguistics that the area of greatest diversity in a language family is likely to be the homeland area of the family (Sapir 1916, Dyen 1956). By this principle, the southern part of the Taiwan highlands appears to be the oldest Austronesian homeland area that can be established by the comparative method, as Puyuma, Tsou and Rukai are apparent first-order subgroups, and the area also contains the rather diverse Nuclear Austronesian languages Paiwan, Amis, Bunun and Kanakanavu-Saaroa. Whether Amis was present in the area before a relatively recent date is not clear, as its villages are scattered along the coast in otherwise Paiwan territory. What is clear, however, is that Bunun has expanded, apparently at the expense of Kanakanavu and Saaroa, and now intrudes between Rukai and the so-called Tsouic languages, which, as suggested by some of the evidence reviewed in this paper, were once in contact (§5.2.2-4).

References

- Adelaar, K. Alexander. 2011. *Siraya: Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language*. Berlin & Boston: De Gruyter Mouton.
- Aikhenvald, Alexandra Y. 2002. *Language Contact in Amazonia*. Oxford & New York: Oxford University Press.
- Bellwood, Peter, Geoffrey Chambers, Malcolm Ross, and Hsiao-chun Hung. 2011. Are ‘cultures’ inherited? Multidisciplinary perspectives on the origins and migrations of Austronesian-speaking peoples prior to 1000 BC. *Investigating Archaeological Cultures: Material Culture, Variability and Transmission*, ed. by Benjamin W. Roberts & Marc Vander Linden, 321-354. New York: Springer.
- van den Berg, René, and Brenda H. Boerger. 2011. A Proto-Oceanic passive? Evidence from Bola and Natügu. *Oceanic Linguistics* 50.1:221-246.
- Blevins, Juliette. 2004. *Evolutionary Phonology*. Cambridge & New York: Cambridge University Press.
- Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai‘i Working Papers in Linguistics* 9.2:1-15.
- Blust, Robert A. 1978. Eastern Malayo-Polynesian: a subgrouping argument. *Second International Conference on Austronesian Linguistics: Proceedings*, ed. by Stephen A. Wurm & Lois Carrington, 181-234. Canberra: Pacific Linguistics.
- Blust, Robert A. 1993. *S metathesis and the Formosan/Malayo-Polynesian language boundary. *Language—A Doorway Between Human Cultures: Tributes to Dr. Otto Chr. Dahl on His Ninetieth Birthday*, ed. by Øyvind Dahl, 178-183. Oslo: Novus Forlag.
- Blust, Robert A. 1995a. The position of the Formosan languages: method and theory in Austronesian comparative linguistics. *Austronesian Studies Relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng, 585-650. Taipei: Institute of History and Philology, Academia Sinica.
- Blust, Robert A. 1995b. Sibilant assimilation in Formosan languages and the Proto-Austronesian word for ‘nine’: a discourse on method. *Oceanic Linguistics* 34.2: 443-453.
- Blust, Robert A. 1996. Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35.2:272-294.
- Blust, Robert A. 1997. Rukai stress revisited. *Oceanic Linguistics* 36.2:398-403.
- Blust, Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 31-94. Taipei: Institute of Linguistics, Academia Sinica.

- Blust, Robert A. 2000. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages. *Time Depth in Historical Linguistics*, ed. by Colin Renfrew, April McMahon & Larry Trask, 311-331. Cambridge: McDonald Institute for Archaeological Research.
- Blust, Robert A. 2001. Squib: Malayo-Polynesian: new stones in the wall. *Oceanic Linguistics* 40.1:151-155.
- Blust, Robert A. 2003a. *Thao Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert A. 2003b. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42.2:438-478.
- Blust, Robert A. 2004. Austronesian nasal substitution: a survey. *Oceanic Linguistics* 43.1:73-148.
- Blust, Robert A. 2005. Review of Lynch, Ross and Crowley 2002. *Oceanic Linguistics* 44.2:544-558.
- Blust, Robert A. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24.1:1-36.
- Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.
- Bybee, Joan L. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14.3: 261-290.
- Cauquelin, Josiane. 1991. The Puyuma language. *Bijdragen tot de Taal-, Land- en Volkenkunde* 147:17-60.
- Chang, Anna Hsiou-chuan. 2006. *A Reference Grammar of Paiwan*. Canberra: The Australian National University dissertation.
- Chang, Chih-Wei Hugh. 2008. *Causative Constructions in Tsou*. Chiayi: National Chung Cheng University MA thesis.
- Chang, Henry Y. 2006. Rethinking the Tsouic subgroup hypothesis: a morphosyntactic perspective. *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on his 70th Birthday*, ed. by Henry Y. Chang, Lillian M. Huang & Dah-an Ho, 565-583. Taipei: Institute of Linguistics, Academia Sinica.
- Chen, Matthew Y., and William S-Y. Wang. 1975. Sound change: actuation and implementation. *Language* 51.2:255-281.
- Comrie, Bernard, and Sandra A. Thompson. 1985. Lexical nominalization. *Language Typology and Syntactic Description*, Vol. 3: *Grammatical Categories and the Lexicon*, ed. by Timothy Shopen, 349-398. Cambridge & New York: Cambridge University Press.
- Dahl, Otto Chr. 1973. *Proto-Austronesian*. Lund: Studentlitteratur.
- Dahl, Otto Chr. 1976. *Proto-Austronesian* (2nd revised edition). Lund: Studentlitteratur.

- Dahl, Otto Chr. 1981. *Early Phonetic and Phonemic Changes in Austronesian*. Oslo: Institute for Comparative Research in Human Culture.
- Dempwolff, Otto. 1937. *Vergleichende Lautlehre des austronesischen Wortschatzes*, Vol. 2: *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen*. Berlin: Dietrich Reimer.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473:79-82.
- Dyen, Isidore. 1956. Language distribution and migration theory. *Language* 32.4:611-626.
- Dyen, Isidore. 1963. The position of the Malayopolynesian languages of Formosa. *Asian Perspectives* 7.1-2:261-271.
- Egerod, Søren. 1965. Verb inflexion in Atayal. *Lingua* 15:251-282.
- Ferrell, Raleigh J. 1969. *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. Taipei: Institute of Ethnology, Academia Sinica.
- François, Alexandre. 2011. Social ecology and language history in the northern Vanuatu linkage: a tale of divergence and convergence. *Journal of Historical Linguistics* 1.2:175-246.
- Geraghty, Paul A. 1983. *The History of the Fijian Languages*. Honolulu: University of Hawai'i Press.
- Geraghty, Paul A., and Andrew Pawley. 1981. The relative chronology of some innovations in the Fijian languages. *Studies in Pacific Languages and Cultures in Honour of Bruce Biggs*, ed. by Jim Hollyman & Andrew Pawley, 159-178. Auckland: Linguistic Society of New Zealand.
- Harvey, Mark. 1982. Subgroups in Austronesian. *Papers from the Third International Conference on Austronesian Linguistics*, Vol. 2: *Tracking the Travellers*, ed. by Amran Halim, Lois Carrington & Stephen A. Wurm, 47-99. Canberra: Pacific Linguistics.
- Ho, Dah-an. 1998. Taiwan Nandaoyu de yuyan guanxi [Genetic relationships among the Formosan languages]. *Chinese Studies* 16.2:141-171.
- Huang, Lillian M. 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364-390.
- Huang, Lillian M. 2012. Nominalization in Sqliq Atayal. Paper presented at the 12th International Conference on Austronesian Linguistics (12-ICAL), July 2-6, 2012. Denpasar, Indonesia.
- Huang, Shuanfan, and Huei-ju Huang. 2005. Causative and applicative: their split and syncretism in Formosan languages. Paper presented at the Taiwan-Japan Joint Workshop on Austronesian Languages, June 24-25, 2005. Taipei: National Taiwan University.

- Imbault-Huart, Camille. 1893. *L'île Formose*. Paris: Ernest Leroux.
- Kaufman, Daniel. 2009. On *pa-, *pa<R> and *pa<ŋ>. Paper presented at the 11th International Conference on Austronesian Linguistics (11-ICAL), June 22-26, 2009. Aussois, France.
- Labov, William. 2010. *Principles of Linguistic Change*, Vol. 3: *Cognitive and Cultural Factors*. Malden: Wiley-Blackwell.
- Li, Paul Jen-kuei. 1972. On comparative Tsou. *Bulletin of the Institute of History and Philology Academia Sinica* 44.2:311-338.
- Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology Academia Sinica* 48.1:1-92.
- Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology Academia Sinica* 52.2:235-301.
- Li, Paul Jen-kuei. 1985a. Linguistic criteria for classifying the Atayalic dialect groups. *Bulletin of the Institute of History and Philology Academia Sinica* 56.4:699-718.
- Li, Paul Jen-kuei. 1985b. The position of Atayal in the Austronesian family. *Austronesian Linguistics at the 15th Pacific Science Congress*, ed. by Andrew Pawley & Lois Carrington, 257-280. Canberra: Pacific Linguistics.
- Li, Paul Jen-kuei. 1995. Formosan vs. non-Formosan features in some Austronesian languages in Taiwan. *Austronesian Studies Relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng, 651-681. Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 1996. *Yilanxian Nandao Minzu yu Yuyan* [*The Austronesian Ethnic Groups and Languages of Yilan County*]. Yilan: Yilan County Government.
- Li, Paul Jen-kuei. 1999. Some problems in the Basay language. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 637-665. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei. 2001. Basaiyu de diwei [The linguistic position of Basay]. *Language and Linguistics* 2.2:155-171.
- Li, Paul Jen-kuei. 2006. Numerals in Formosan languages. *Oceanic Linguistics* 45.1: 133-152.
- Li, Paul Jen-kuei, and Shigeru Tsuchida. 2001. *Pazih Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei, and Shigeru Tsuchida. 2009. Yet more Proto Austronesian infixes. *Discovering History through Language: Papers in Honour of Malcolm Ross*, ed. by Bethwyn Evans, 345-362. Canberra: Pacific Linguistics.
- Liao, Hsiu-chuan. 2011a. Some morphosyntactic differences between Formosan and Philippine languages. *Language and Linguistics* 12.4:845-876.

- Liao, Hsiu-chuan. 2011b. On the development of comitative verbs in Philippine languages. *Language and Linguistics* 12.1:205-237.
- Lynch, John, Malcolm Ross, and Terry Crowley. (eds.) 2002. *The Oceanic Languages*. Richmond, Surrey: Curzon.
- Pawley, Andrew. 1996. On the position of Rotuman. *Reconstruction, Classification, Description: Festschrift in Honor of Isidore Dyen*, ed. by Bernd Nothofer, 85-120. Hamburg: Abera.
- Pawley, Andrew. 1999. Chasing rainbows: implications of the rapid dispersal of Austronesian languages for subgrouping and reconstruction. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 95-138. Taipei: Institute of Linguistics, Academia Sinica.
- Pawley, Andrew. 2006. Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 115.3:215-258.
- Pawley, Andrew, and Roger C. Green. 1984. The Proto-Oceanic language community. *Journal of Pacific History* 19.3:123-146.
- Pejros, Ilia. 1994. Some problems of Austronesian accent and *t~*C (notes of an outsider). *Oceanic Linguistics* 33.1:105-127.
- Post, Ursula. 1992. *Binukid Dictionary*. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.
- Reid, Lawrence A. 1966. *An Ivatan Syntax*. Honolulu: University of Hawai'i Press.
- Ringe, Don, Tandy Warnow, and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100.1:59-129.
- Ross, Malcolm. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Canberra: Pacific Linguistics.
- Ross, Malcolm. 1992. The sound of Proto-Austronesian: an outsider's view of the Formosan evidence. *Oceanic Linguistics* 31.1:23-64.
- Ross, Malcolm. 1995a. Reconstructing Proto Austronesian verbal morphology: evidence from Taiwan. *Austronesian Studies Relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng, 727-791. Taipei: Institute of History and Philology, Academia Sinica.
- Ross, Malcolm. 1995b. Some current issues in Austronesian linguistics. *Comparative Austronesian Dictionary*, Vol. 1, ed. by Darrell T. Tryon, 45-120. Berlin & New York: Mouton de Gruyter.
- Ross, Malcolm. 1997. Social networks and kinds of speech-community event. *Archaeology and Language*, Vol. 1: *Theoretical and Methodological Orientations*, ed. by Roger M. Blench & Matthew Spriggs, 209-261. London: Routledge.
- Ross, Malcolm. 1998. Proto Oceanic phonology and morphology. *The Lexicon of Proto Oceanic: The Culture and Environment of Ancestral Oceanic Society*, Vol. 1:

- Material Culture*, ed. by Malcolm Ross, Andrew Pawley & Meredith Osmond, 15-35. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002a. The history and transitivity of western Austronesian voice and voice-marking. *The History and Typology of Western Austronesian Voice Systems*, ed. by Fay Wouk & Malcolm Ross, 17-62. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2002b. Final words: research themes in the history and typology of western Austronesian languages. *The History and Typology of Western Austronesian Voice Systems*, ed. by Fay Wouk & Malcolm Ross, 451-474. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on his 70th Birthday*, ed. by Henry Y. Chang, Lillian M. Huang & Dah-an Ho, 521-563. Taipei: Institute of Linguistics, Academia Sinica.
- Ross, Malcolm. 2007. Calquing and metatypy. *Journal of Language Contact: Thema* 1:116-143.
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, ed. by K. Alexander Adelaar & Andrew Pawley, 295-326. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2011. Some Proto Austronesian coronals reexamined. Manuscript. Canberra: The Australian National University.
- Ross, Malcolm. 2012. Reconstructing Proto Austronesian verb classes. Paper presented at the 19th Annual Meeting of the Austronesian Formal Linguistics Association (AFLA-19), June 26-30, 2012. Taipei: Academia Sinica.
- Ross, Malcolm. (forthcoming). Diagnosing contact processes from their outcomes: the importance of life stages. *Journal of Language Contact*.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43.2:411-444.
- Sagart, Laurent. 2010. Is Puyuma a primary branch of Austronesian? *Oceanic Linguistics* 49.1:194-204.
- Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.3:560-588.
- Sapir, Edward. 1916. *Time Perspective in Aboriginal American Culture*. Ottawa: Government Printing Bureau.
- Starosta, Stanley. 1974. Causative verbs in Formosan languages. *Oceanic Linguistics* 13.1-2:279-369.
- Starosta, Stanley. 1985. Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou. *Austronesian Linguistics at the 15th Pacific Science Congress*, ed. by Andrew Pawley & Lois Carrington, 281-312. Canberra: Pacific Linguistics.

- Starosta, Stanley. 1996. The position of Saaroa in the grammatical subgrouping of Formosan languages. *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Languages and Linguistics*, Vol. 3, ed. by Suwilai Premisrat, 944-966. Salaya: Institute of Language and Culture for Rural Development, Mahidol University at Salaya.
- Starosta, Stanley. 2002. The rise and fall and rise and fall of Proto Malayo-Polynesian. *Collected Papers on Southeast Asian and Pacific Languages*, ed. by Robert S. Bauer, 185-203. Canberra: Pacific Linguistics.
- Starosta, Stanley, Andrew Pawley, and Lawrence A. Reid. 1981. The evolution of focus in Austronesian. Paper presented at the 3rd International Conference on Austronesian Linguistics, January 24-29, 1981. Denpasar, Bali, Indonesia.
- Starosta, Stanley, Andrew Pawley, and Lawrence A. Reid. 1982. The evolution of focus in Austronesian. *Papers from the Third International Conference on Austronesian Linguistics*, Vol. 2: *Tracking the Travellers*, ed. by Amran Halim, Lois Carrington & Stephen A. Wurm, 145-170. Canberra: Pacific Linguistics.
- Starosta, Stanley, Andrew Pawley, and Lawrence A. Reid. 2009. The evolution of focus in Austronesian. *Formosan Linguistics: Stanley Starosta's Contributions*, Vol. 2, ed. by Elizabeth Zeitoun, 329-480. Taipei: Institute of Linguistics, Academia Sinica.
- Szakos, József. 1994. *Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan*. Bonn: Rheinische Friedrich-Wilhelms-Universität dissertation.
- Szakos, József. (n.d.) Cou-Wörterbuch. Unpublished computer file. Taichung: Providence University.
- Teng, Stacy Fang-ching. 2008. *A Reference Grammar of Puyuma, an Austronesian Language of Taiwan*. Canberra: Pacific Linguistics.
- Teng, Stacy Fang-ching, and Malcolm Ross. 2010. Is Puyuma a Primary Branch of Austronesian? A reply to Sagart. *Oceanic Linguistics* 49.2:543-558.
- Tung, T'ung-ho. 1964. *A Descriptive Study of the Tsou Language*. Taipei: Institute of History and Philology, Academia Sinica.
- Thomason, Sarah Grey, and Terrence S. Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Thomson, John. 1873. Notes of a journey in southern Formosa. *Journal of the Royal Geographical Society* 43:97-107.
- Thomson, John. 1877. *Dix ans de voyages dans la Chine et l'Indo-Chine*. Paris: Hachette.
- Trudgill, Peter. 1974. Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. *Language in Society* 3.2:215-246.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.

- Tsuchida, Shigeru. 1980. Puyuma-go (Tamarakao hōgen) goi: fu gohōgaisetsu oyobi tekisuto [Puyuma (Tamalakaw dialect) vocabulary: with grammatical notes and texts]. *Kuroshio no Minzoku, Bunka, Gengo* [Ethnology, Cultures and Languages along the Black Current], ed. by Kuroshio Bunka no Kai [Black Current Cultures Committee], 183-307. Tokyo: Kadokawa Shoten.
- Tsuchida, Shigeru. 1981. Unpublished wordlists of Amis dialects collected in 1980-81. Tokyo: University of Tokyo.
- Tsuchida, Shigeru. 1982. *A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan*, Part I: *West Taiwan*. Tokyo: Faculty of Letters, University of Tokyo.
- Tsuchida, Shigeru, and Yukihiro Yamada. 1991. Ogawa's Siraya/Makatao/Taivoan. *Linguistic Materials of the Formosan Sinicized Populations*, Vol. 1: *Siraya and Basai*, ed. by Shigeru Tsuchida, Yukihiro Yamada & Tsunekazu Moriguchi, 1-194. Tokyo: Department of Linguistics, University of Tokyo.
- Van Der Mark, Sheena. 2007. *A Grammar of Vinitiri: An Austronesian Language of Papua New Guinea*. Melbourne: La Trobe University dissertation.
- Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
- Winter, Bodo. 2010. A note on the higher phylogeny of Austronesian. *Oceanic Linguistics* 49.1:282-287.
- Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. *Parangal kay Cecilio Lopez*, ed. by Andrew B. Gonzalez, 71-91. Quezon City: Linguistic Society of the Philippines.
- Wolff, John U. 1991. The Proto-Austronesian phoneme *t and the grouping of the Austronesian languages. *Currents in Pacific Linguistics: Papers on Austronesian Languages and Ethnolinguistics in Honour of George W. Grace*, ed. by Robert A. Blust, 535-549. Canberra: Pacific Linguistics.
- Wolff, John U. 1997. The Proto-Austronesian voiced apical and palatal stops *d and *j. *Proceedings of the Seventh International Conference on Austronesian Linguistics*, ed. by Cecilia Odé & Wim Stokhof, 581-601. Leiden: Dept. of Languages and Cultures of South East Asia and Oceania, Leiden University.
- Wolff, John U. 2007. Do the Malayo-Polynesian languages constitute a subgroup of the Austronesian language family? *Studies in Philippine Languages and Cultures* 15:13-24.
- Wright, Richard A. 1996. *Consonant Clusters and Cue Preservation in Tsou*. Los Angeles: University of California dissertation.
- Wu, Chunming. 2006. Verb serialization in Kanakanavu. *University System of Taiwan Working Papers in Linguistics* 2:109-139.

- Zeitoun, Elizabeth. 2000. Dymanic vs. stative verbs in Mantauran (Rukai). *Oceanic Linguistics* 39.2:415-427.
- Zeitoun, Elizabeth. 2002. Reciprocals in the Formosan languages: a preliminary study. Paper presented at the Ninth International Conference on Austronesian Linguistics (ICAL-9), January 8-11, 2002. Canberra: The Australian National University.
- Zeitoun, Elizabeth. 2005. Tsou. *The Austronesian Languages of Asia and Madagascar*, ed. by K. Alexander Adelaar & Nikolaus P. Himmelmann, 259-290. London & New York: Routledge.
- Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.
- Zeitoun, Elizabeth. (ed.) 2009. *Formosan Linguistics: Stanley Starosta's Contributions*, Vol. 2. Taipei: Institute of Linguistics, Academia Sinica.
- Zeitoun, Elizabeth, and Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391-414.
- Zeitoun, Elizabeth, and Stacy F. Teng. 2009. From ki-N 'get N' in Formosan languages to ki-V 'get V-ed' (passive) in Rukai, Paiwan and Puyuma. *Discovering History through Language: Papers in Honour of Malcolm Ross*, ed. by Bethwyn Evans, 479-500. Canberra: Pacific Linguistics.
- Zorc, R. David. 1986. The genetic relationships of Philippine languages. *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. by Paul A. Geraghty, Lois Carrington & Stephen A. Wurm, 147-173. Canberra: Pacific Linguistics.

[Received 9 April 2012; revised 24 June 2012; accepted 6 August 2012]

School of Culture, History & Language
The Australian National University
Canberra, ACT 0200
Australia
Malcolm.Ross@anu.edu.au

核心南島語假設之辯證

Malcolm Ross

澳洲國立大學

Ross (2009) 曾提出「核心南島語假設」，在此假設下，卑南語、鄒語以及魯凱語各自為南島語的第一支分群，而所有其他的南島語則同屬於另一個第一支分群，稱之為「核心南島語」。屬於「核心南島語」這一分群的語言有一個複雜的共同創新，即在古南島語作為名物化的一些標記，在這些語言變成動詞的標記。本文分為兩部分。第一部分探討歷史語言學家在做分群時所使用的不同種類證據，並檢視支持「核心南島語假設」的證據及其他近年關於南島語分群假設的證據，依此評估這些不同假設的可信度。第二部分則討論與「核心南島語假設」不相容的「鄒語群假設」，作者認為儘管「鄒語群假設」為許多人所接受，然而並沒有證據支持此一假設。相反地，所謂「鄒語群」所反映的是長久以來鄒語和卡那卡那富語及沙阿魯阿語因為接觸而起的現象。結論認為，台灣南部山區也許就是南島語的原鄉。

關鍵詞：南島語，核心南島語，台灣南島語，鄒語群，鄒語，分群